AGENTS FOR THE SALE OF MADRAS GOVERNMENT PUBLICATIONS.

IN INDIA.

BUTTERWORTH & Co. (LTD.), 6, Hastings Street, Calcutta.
R. CAMBRAY & Co., Calcutta.
E. M. GOPALAKRISHNA KONE, Pudumantapam, Madura.
HARTLEYS, Mount Road, Madras.
HIGGINBOTHAMS (LTD.), Mount Road, Madras.
V. KALYANARAMA IYER & Co., Esplanade, Madras.
G. C. LOGANATHAM BROTHERS, Madras.
S. MURTHY & Co., Madras.
G. A. NATESAN & Co., Madras.
The Superintendent, NAZAIR KANUN HIND PRESS, Allahabad.
NIVASARKAR, Manager, "Hitawada," Nagpur.
P. R. RAMA IYER & Co., Madras.
RAMAKRISHNA & SONS, Lahore.
R. SUNDER PANDURANG, Kalbadevi Road, Bombay.
D. B. TARAPOREVALA SONS & Co., Bombay.
THACKER & Co. (LTD.), Bombay.
THACKER, SPINK & Co., 3, Esplanade East, Calcutta
S. VAS & Co., Madras.
S.P.C.K. PRESS, Vepery, Madras.

IN THE UNITED KINGDOM.

B. H. BLACKWELL, 50 and 51, Broad Street, Oxford.
CONSTABLE & Co., 10, Orange Street, Leicester Square, London, W.C.
DEIGHTON, BELL & Co. (LTD.), Cambridge.
EAST AND WEST (LTD.), 3, Victoria Street, London, S.W. 1.
T. FISHER UNWIN (LTD.), 1, Adelphi Terrace, London, W.C.
GRINDLAY & Co., 54, Parliament Street, London, S.W.
KEGAN PAUL, TRENCH, TRÜBNER & Co. (LTD.), 68—74, Carter Lane, London, E.C., and 39, New Oxford Street, London, W.C.
HENRY S. KING & Co., 65, Cornhill, London, E.C.
P. S. KING & SON, 2 and 4, Great Smith Street, Westminster, London, S.W.
LUZAC & Co., 46, Great Russell Street, London, W.C.
B. QUARITCH, 11, Grafton Street, New Bond Street, London, W.
W. THACKER & Co., 2, Creed Lane, London, E.C.
OLIVER AND BOYD, Tweeddale Court, Edinburgh.
E. PONSONBY (LTD.), 116, Grafton Street, Dublin.
WHELDON & WESLEY (LTD.), 2, 3 & 4, Arthur Street, New Oxford Street, London, W.C. 2

ON THE CONTINENT.

ERNEST LEROUX, 28, Rue Bonaparte, Paris.
FRIEDLANDER AND SOHN, Berlin.
MARTINUS NIJHOFF, The Hague, Holland.
OTTO HARRASSOWITZ, Leipzig.

University of Madras

PSYCHOLOGICAL TESTS OF MENTAL ABILITIES

BY

A. S. WOODBURNE, M.A., PH.D.

Professor of Psychology, The Madras Christian College; Author of "The Relation between Religion and Science"

MADRAS

PRINTED BY THE SUPERINTENDENT, GOVERNMENT PRESS

PRICE, 2 *rupees* 8 *annas*] 1924

PREFACE.

At a meeting of the Central Advisory Board of Education for India in October 1921 the subject of intelligence tests was discussed. As one result of the discussion the Principals of the Teachers' Colleges at Saidapet, Jubbulpore, Lahore, and Dacca were asked to conduct experiments with the children attending their model schools, using the Stanford-Binet tests. The results of these experiments in Saidapet are given in Bulletin No. 15 of the Teachers' College. The results for all the experiments have been made the basis of a provisional set of tests which, it is hoped, will be suitable for Indian schools, and which will be published shortly by the Bureau of Education.

Those who were in charge of the experiments in Saidapet felt the need of more information on the subject. In response to their request for such a course, the Syndicate of the University of Madras at the request of the Board of Studies in Teaching invited the present writer to deliver a course of ten University lectures on the subject. Accordingly ten lectures were given at the Museum Theatre during the months of December 1922 and January 1923. The present volume is the text of the lectures with a few alterations, but much additional matter which the limits of time prevented the writer from presenting in the lectures. It is hoped that, in their printed form, the lectures may serve an even larger purpose in informing the teachers of South India of this most important phase of educational psychology, and of inspiring greater practical effort in the field.

The purpose of the lectures was purely to impart information. The debt which the author owes to many outstanding scholars is evident on almost every page. He has endeavoured to give sufficient information in the footnotes

to enable interested students to know how to secure some o the best available books and journals. The final chapter is an attempt to summarize the more important problems that face us in India who are trying to make any practical use of mental tests in our educational work. If this pioneer in the literature on psychological tests in India serves to promote further discussion and experimentation, it will have served its purpose.

I have to acknowledge with thanks the work done by Mr. Paul Lawrence, B.A. (Honours), in compiling the index.

MADRAS, *November* 1923.

A. S. WOODBURNE.

CONTENTS.

PSYCHOLOGICAL TESTS OF MENTAL ABILITIES.

CHAPTER I.

HISTORICAL DEVELOPMENT.

One of the most noteworthy developments of educational psychology in recent times is the development of standardized tests for the measurement of mental abilities. Until comparatively recent times there was no attempt made to attain any unity of method in the judgement of mental facts, and indeed the great majority of people have very little conception of standardized measurements, even yet. The lamentable fact is not that the mass of the people are deficient in knowledge of this type, but that many of those whose business it is to make tests of mental abilities have not the technique for making their tests according to any uniform scale. But the past eighteen years have been the beginning of a new era in this direction. Not only has a prodigious amount of labour been expended on this problem by educational psychologists, but even now a small army of investigators are working on various phases of the problem.

The name with which the beginnings of the work of standardization in testing is imperishably associated is that of Alfred Binet. The Board of Education commissioned him to investigate the problem of feeble-mindedness in the Parisian public schools, and it was as an instrument to help him in that task that he worked out his first scale of tests. Binet was an indefatigable worker, and was constantly engaged in the task of revision and experiment from the time that he began the work until his premature death in 1913. His first scale was issued in 1905, a revision was published in 1908, a second revision in 1911, and when he died he was working on a further revision.

We need carefully to distinguish the work of Binet and those who carried on the work which he began from previous work done in mental measurement. To say that Binet began the work of standardizing tests is not equivalent to saying that he began the measurement of mental abilities. For the attempt to measure mental ability goes back a very long time. It is a known historical fact that China had a system of competitive examinations in vogue 4,000 years ago. Not only so, but practically all cultures give evidence of attempts to test mental ability as something distinct

from physical power. Ballard refers to the riddle as an example of this tendency.[1] He cites the instances of Oedipus and the Sphinx, and Samsom and the lion. We might add the story which is to be found in varying forms in Hebrew and Hindu folk-lore, the story of the two women who came before the king, each laying claim to be the mother of the same child, and the king's test of the real mother by suggesting the cutting of the child in two and its partition between the two. But, of course, the riddle is not to be interpreted as a careful measurement of intelligence, although it has served as a mental test on occasion.

Some attempts have been made to measure mentality on the basis of its physical concomitants. At bottom, they assume a kind of psycho-physical parallelism, though we must not confuse them with the specific movement which is known by that designation. The most notable attempt is that which is known variously as "Cranioscopy," "Physiognomy," or "Phrenology." It is an attempt to correlate mental abilities and faculties with cerebral localities and configurations. Phrenology has proved to be a scientific absurdity both on the side of its physiological and its psychological assumptions. In particular the actual localization of specific cerebral functions which has been accomplished has completely negativized the assumptions of the phrenologists. But the interest for us here is the fact of an honest attempt to find a basis for measuring mental facts.

Other attempts have been made to find a basis for measuring mental ability in physical or physiological facts. One is that of the Italian criminologist and psychiatrist, Lombroso. He believed, as Auguste Comte had taught, that mental facts are all referrable to biological causes. The net result of his investigations was the theory that criminals possess a greater average number of mental, neural and physical abnormalities than do the non-criminals. Though his theory of a "criminal type" has been severely criticized, yet it is admitted to contain a modicum of truth.

Sir Francis Galton, the celebrated English anthropologist, approached the problem from the angle of his special interest. He was concerned with the question of possible means for improving the human race, which included the eugenic problem. He wanted to find new ways of gaining social control for the improvement of racial qualities both physical and mental. For he was convinced that there was some degree of correspondence between mental abilities and certain physiological factors such as the character of finger prints. But the subsequent investigations of Karl Pearson and others have indicated that there is very little

[1] Mental Tests, p. 3.

correlation between these physiological facts and mentality Neither subnormality nor genius are facts that we can discover by means of physical measurements or contour.

At the same time, these facts do not prove that there is no association between physical and mental facts. To be sure, the relation between body and mind is one of the persistent problems of psychology, and indeed of metaphysics. Various solutions have been propounded, but the issue in our day has been narrowed down to the alternative of parallelism or interaction. One of the facts which we are prone to neglect in the discussion is the fact that the dissociation of mind and body is one that has been made in the interests of our scientific inquiries and that in experience we do not experience them apart at all. We have to deal in actual life with a psycho-physical organism which is a unity. Consequently we have a right to expect some form of interaction or parallelism between these phases or functions of life which, for theoretical interests, we have separated.

One of the main differences between the psychology of the past generation and that of to-day is that the earlier was structural and static, whereas the more recent is functional and dynamic. I can think of no more apt illustration than that of the difference between the stomach as an organ, and digestion as a process. It is, in other words, the difference that subsists between anatomy and physiology. Now the reason for the failure of the physiognomists and phrenologists, as well as of Lombroso and Galton, was that they were working on the old structural basis. Let it be once recognized that we are dealing with living processes pertaining to a unified organism and the problem takes on new significance. There is a profound truth, which perhaps the behaviourists are liable to over-emphasize, in the unity of our behaviour. We cannot study the physical and the mental phases of conduct as factors distinct and disparate. So that any attempt to measure mental abilities must take cognizance of this fundamental unity. It is true that there are forms of motor ability which do not carry as necessary concomitants a marked intellectual ability. This may be illustrated in the tapping experiment which tests the number of taps per second which a person can make with a pencil on paper, and which may have its use as a measurement of the relation between motor ability of a certain sort and fatigue, but does not demand any great mental skill. At the same time, Whipple in summing up the results of the experiment says that there is a positive correlation between tapping ability and mental ability on the one hand and social status on the other, and that cases of epilepsy, insanity and retardation show a corresponding inability in tapping.

The reaction-time experiments mark a further stage in the development of mental tests. In this experiment the subject is required to respond by some motor response to a given signal for which he is warned to be prepared, and the time required to respond to the stimulus is noted. The experiment is varied in respect to the different end organs of sense receiving the stimuli, and in regard to the complexity of the response required, whether simple, alternative or associative. These responses are found to correlate very closely to a number of important practical situations. The motor-man on the tram-car applying the brakes on the signal of the conductor or guard, the athlete's response to signals on the field of sport, and the boxer's dodging the blows of his opponent are illustrations in point. On the other hand in the laboratory experiments the subject is usually better prepared than in actual situations, so that reaction-time in ordinary life is a bit longer than in the tests.

The next attempt to measure psychical factors is seen in the study of the relationship between stimuli and sensations. One of the pioneers in this field was Weber (1795—1878) whose investigations led him to the conclusion that the least addition to a stimulus caused a difference in the intensity of the sensation on which basis he worked out a system of gradations, showing the relations which obtain among sensation-intensities as we perceive them. Fechner carried the implication of Weber's hypothesis to its logical conclusion and showed that the sensation varies with the stimulus, even when the difference is too small to be perceptible or measureable. Subsequent investigation has verified the general conclusion of Weber in regard to the relation between stimulus and sensation though it has also made it evident that it cannot be so accurately determined with mathematical precision. These experiments mark the beginnings of the laboratory method in psychology, and their interest for us in this connection lies in the fact that they indicate a disposition to measure psychical factors in experience.

But the work which was begun by Binet is of the more specialized type which interests us just now. It was not concerned with the correlation of physical and mental abilities, nor yet with the measurement of any specific ability, but with intelligence in the large. The immediate problem was that of feeble-mindedness, or perhaps it would be better to say retardation in the public schools, for the designation feeble-mindedness was not yet much in use. The school authorities in Paris were confronted with the fact that there were a great many children who, for reasons which they could not adequately understand, were backward in their school work. There were cases where the children did not seem to be able to attend as they should to their teachers. Others showed evidence of constitutional moral difficulties. Still others simply

could not learn. So Binet was asked to make a study of the problem with the hope that remedial measures might be devised. Fortunately Binet and his collaborator, Simon, decided to strike out on new and independent lines, rather than to follow any of the paths to which we have already referred. If they had tried merely to revise phrenology or to devise some new experiments in reaction-time or in motor ability, the probability is that we would still be deficient in standardized tests of mental abilities. Moreover these investigators realized that it was not achieved knowledge which they were required to test. That was already being tested in a way by the public examinations. But the particular information for which they sought was the reason for the backwardness of those who had failed to attain the average amount of knowledge expected of children of their age and school opportunities.

For that purpose Binet devised two scales, each carefully graded. The first scale was intended as a loose device whereby it would be possible to separate without delay those about whose intelligence there could be no doubt from those who were suspected of mental defectiveness. The second scale was devised for more accurate measurement, and aimed to give a final criterion of mental deficiency. Binet realized from the outset that no one test could be accepted as adequate because it might only test one phase of intelligence in which the subject might or might not be proficient. He decided that it would be better to provide several brief tests, thus giving the child abundance of opportunity to show his ability and accomplishments. The scales were graduated so as to test the intelligence of children at various ages, from three years onward. Before fixing the scale with any definiteness it was necessary to ascertain what tests would be appropriate to the various ages. In other words, as Ballard has put it, "before testing children with a test, he first tested the test with children."[1] By testing a large number of children he was enabled to discover the lowest age at which a child was able to pass that test. If 75 per cent of the children of a certain physical age were able to pass the test correctly he then fixed upon it as a test for that age. To be sure subsequent applications led him to make revisions, as larger groups affected the averages. But he was always ready to make such adjustments, and indeed adjustments and revisions are always being made as more data comes to hand from the practical application of the tests.

The measurement of intelligence in Binet's system is on the basis of mental age. The average child is taken as the criterion. An average child of any particular age, say ten years and three months, is regarded as having a *mental age* of that particular age.

[1] Mental Tests, p. 35.

So that we have a technique for determining a child's mental age regardless of what may be the child's physical age. If 75,000 out of 100,000 ten-year old children pass the tests for that age, then we can fix a child's mental age at ten who is successful with those tests but is not able to pass any higher tests. On the other hand if a child of seven is successful with the ten-year old tests, we say that the child's mental age is ten. If a child's mental age is the same as his chronological age, the child is of average intelligence, neither dull nor bright. If the mental age is decidedly below the chronological age, he is sub-normal; if it is above, he is superior. The value consists not simply in the ability to find out what children are dull and who are bright, but also the degree of their inferiority or superiority in terms of years retarded or advanced.

The relation between these two ages, chronological and mental, is then the measurement of the child's intelligence. It is expressed by what is called the *Intelligence Quotient*, a term which was invented by the German psychologist, Stern, and which has been employed so commonly that it has been conveniently abbreviated to "I.Q." The I.Q. of a child is ascertained by the percentage plan of dividing the mental age by the chronological. For example, a child whose mental and physical ages correspond has just 100 per cent intelligence, or in other words has an I.Q. of 100. If the mental age were 12, while the chronological age were 8, the I.Q. would be $\frac{12}{8} \times 100 = 150$. A child of 6 years' mental age and 8 years' physical age would have an I.Q. of $\frac{6}{8} \times 100 = 75$.

The Binet scale has given in this way a new basis for the classification of mentality. Before his day there was no available technique for that purpose, and we ordinarily spoke of people as in three classes, the sane, the insane and the imbeciles. Now we have come to see that intelligence is a function or group of functions which it is possible to classify into as many classes as we choose on the basis of I.Q. Normality is considered to include the range between 90 and 110 I.Q. The others are divided and subdivided pretty much in accordance with the wishes of the experimenter. There are the border-line cases, the dull normals, and the distinctly feeble-minded, and the latter are again often divided into the high-grade and the low-grade morons. Below these again are, of course, the out and out idiots or imbeciles. On the other side there are those who are superiors, those who are very superior, and at the top the geniuses. The highest extreme thus far ascertained is about 200, whereas the lowest is about zero.

It is not surprising that the Binet scale, being the first in the field, should be the object of a great deal of criticism. Yet, despite the criticisms, the subsequent scales that have been devised have been almost all modelled on the Binet plan, if not acknowledged revisions of it. One of the criticisms preferred is that Binet's test

were too few. He had fifty-four tests in all, which amounted to about five to each year. Since the object is to ascertain what a child can do, the child ought to be given the largest possible opportunity to do justice to himself. So, as we shall see presently, some of the revisers of the scale have added further tests to broaden the scope of the test. Another, and perhaps more serious, criticism was that the Binet tests depend too largely on the use of language. It has been found that there are some children, and some adults too, as the American Army tests show, who are not deficient in intelligence, though circumstances have so contrived as to make them defective in the use of language. That is particularly the case with such tests as involve the use or understanding of abstract terms. It is also a handicap to a child whose native tongue is different to that in which the test is being given, and to a deaf child. Attempts have been made to make good that deficiency in the Binet system by the devising of "performance tests" in which language plays a very insignificant part.

I have already referred to the fact that Binet had two scales. His second scale, which he called his *barème d' instruction*, consisted of questions of the ordinary examination type, carefully selected and standardized. Binet's plan was to conduct a pedagogical examination of acquired knowledge along with his psychological examination of natural intelligence. His motive was the same in the case of the application of the *barème d'instruction* as in the case of the intelligence tests, to find out what children were subnormal. It was based on the average performance of a large number of Parisian children, and purported to supply a ready method of judging a child after testing him in the three branches of reading, arithmetic and spelling. We shall have occasion later to refer to the actual plan of the *barème* and to its trustworthiness as a scale of mental measurement.

In the case of both of the Binet scales, judgement should not be pronounced on their final adequacy, but rather on the discovery which he made of a new method, a new tool which has been subsequently largely developed and is still being improved. The great element in his contribution is his decision that mental abilities must be judged in accordance with a scale rather than on a simple all-or-none performance. It should also be remembered that Binet's aim was not so much the devising of a technique whereby he could grade school children, but the discovery of the sub-normal. The age-performance scale which he devised for that particular purpose has come to be far more useful than he dreamed, and to serve a much broader field in educational psychology than he had hoped. It gives some indication of the wide range of interest and use of the Binet tests when one realizes how much literature has been published on the subject. Whipple explains the absence of any extended reference to them in his *Manual* on

the ground that there have been several handbooks which explain them quite adequately (including Goddard, Kuhlmann, Schwegel, Terman, Town and Winch), and further because the literature is so exhaustive. Kohs' bibliography which was brought up-to-date in June 1914—nine years ago—contained 254 titles.

About the time that Binet published his first scale, i.e., 1908, Mr. Cyril Burt was carrying on a series of experiments with school children at Oxford and later he extended his operations to Liverpool. His plan was to select a group of children of a certain age, and then to procure from their teachers a judgement of their comparative intelligence, the judgement to be based partially on examination results and partially on personal contact. Then twelve psychological tests, varying in range and complexity, were administered. The test of the test was its correlation with the judgement of the teacher. If the correlation was high it was considered to be a satisfactory test of intelligence. Mr. Burt believed moreover that the development of intelligence ought to involve a corresponding expansion of the power to reason, so that as the age advanced he included tests which called for more of the element of reasoning. In this way the judgement of common sense were substantiated by the findings of psychology, the part of psychology being to standardize rather than to invent any new criterions of intelligence. Later on Mr. Burt, in collaboration with Dr. Simon who had also been Binet's collaborator, translated the Binet tests into English. As pointed out, the original tests were standardized on the basis of experiments with the school children of Paris. Mr. Burt carried on the work with the children of the London public schools, which led him to make certain revisions and modifications. The complete set of tests in accordance with the translation and revision of Mr. Burt is published by Dr. Ballard in his book on *Mental Tests* as Chapter IV. Mr. Burt's experience with the Binet tests, like that of many other workers in the field, was that they are much more suited to the junior than to the senior grades. Accordingly he worked out a number of reasoning tests with which to test children of the older grades. The *Journal of Experimental Pedagogy* for June and December 1919 contains Mr. Burt's own account of the tests together with his conclusions on the subject of reasoning in school-children. There are six or seven tests arranged for each year from seven to fourteen, although it was not his intention to use all of the tests in a series on any one child for fear of inducing fatigue. His idea was that the larger number would enable the experimenter to vary the tests or else he could give the remaining tests subsequently. The shorter list contains but seventeen tests, two for each age except the first which has three.

One of the earliest translations and revisions of the Binet scale was made by H. H. Goddard. He published in the *Training School*

Bulletin in 1911, "The Binet-Simon Measuring Scale of Intelligence Revised."[1] There are still a considerable number of workers who make use of the Goddard revision. In the same year (1911) Goddard published also "Two Thousand Children Tested by the Binet Measuring Scale of Intelligence," an investigation the value of which comes out in the definition and discrimination of feeble-mindedness as distinct from normality. This is indeed a question to which Goddard subsequently paid a good deal of attention as witness his publication in 1919 of the "Psychology of the Normal and Sub-normal."[2]

The revision which is probably used the most extensively and known in Britain and in India the best is that made by Prof. Lewis M. Terman and his associates of the Leland Stanford University, and known as the Stanford revision. The revised tests, and the method of valuation, together with a good deal of additional matter of value, is published in Terman's book, *The Measurement of Intelligence*.[3] The scale adheres fairly closely to the original one produced by Binet, its contribution being in the way of additional tests, further standardization, and in the use of the Intelligence Quotient to which reference has already been made. Whereas the Binet scale consisted of an average of five tests for each age, the Stanford revision adds one for each year, as well as one or two alternatives. Yet the general character of the tests which were added was much the same as those in the original scale. The Stanford revisers felt that the mental age was not a sufficiently accurate criterion of mental ability, and chose instead the Intelligence Quotient as proposed by Stern, which was the ratio of mental age to chronological age.

In all cases the child is tested as an individual, and under standardized conditions which, as far as possible, shall be of a nature that will remove all tendencies to shyness or nervousness. The instructions must first of all be given with all the necessary detail and explicitness, though, of course, the examiner must guard against giving them in such a way that by hint or sign the child will be able to find a clue to the problem. Having made sure that the examinee understands what he has to do, the examiner may proceed. It is customary to start with those tests which are assigned to the chronological age just below that of the subject. If the child fails in any of those tests, then the examiner should go back and give the tests of the previous group. The examination should be continued until the child fails in all of the tests, except one. Terman's plan was to include a range starting with the year yielding but one failure and ending with the year having but one

[1] Vol. viii, pp. 56-62.
[2] New York: Dodd, Mead & Co.
[3] Boston: Houghton Mifflin Co, 1916.

success. In estimating the mental age of the child, the examiner should take as a base the year at which the child passes all the tests, and then add one-fifth of a year for every subsequent test passed.

It is quite evident from the material which he gives that Terman, like Binet, was interested in the retarded children. In the beginning of his book he gives us the information that in the United States of America there are from 10 per cent to 15 per cent of school-children retarded by two years, and between 5 and 8 per cent retarded by 3 years. It has been computed that the Government of that country expends a sum equivalent to about Rs. 120 crores annually for the re-education of backward pupils. This, coupled with the psychological study of crime and delinquency, he gives as the *raison d'être* for the interest in the measurement of intelligence. Concerning that matter Terman has summarized a great deal of information which has come to light as a result of the investigations of a number of psychologists and sociologists. Investigations were made in a number of State institutions, reformatories, homes for delinquents, and courts which specially deal with delinquents, both juvenile and adult. In all of these cases it was ascertained that a considerable percentage, ranging from 15 per cent to 50 per cent of the subjects tested were sub-normal mentally. And this result was obtained by psychological examination, in many cases after the subjects had been pronounced to be mentally sound by the examining medical authorities. Terman also quotes from the historical surveys of certain families that owe their place in the hall of fame to the abnormal number of delinquents and criminals which they have been able to include in their numbers. One family—the Hill family—it is reckoned has cost the State of Massachusetts a sum equivalent to Rs. 15 lakhs within the space of 60 years, in addition to the disease and crime which they spread among other families. Investigation revealed the fact that out of 709 members of the family, 48 per cent were feeble-minded, while 24 per cent were criminal, 30 per cent alcoholics, 24 per cent of the women had had illegitimate children, and 10 per cent of them were acknowledged prostitutes. In a similar way, the Juke family has cost the State of New York within the space of 75 years a sum equal to Rs. 40 lakhs, and the Nam family had cost the State a sum equivalent to Rs. 45 lakhs.

One of the most patent uses of a scale of intelligence is to deal with situations such as these. If we are in possession of a technique whereby we can separate the feeble-minded from the normal, we can deal with both classes with greater fairness. Obviously the normal child is held back in a class in school by the sub-normal. The progress which the class makes can be no faster than that of the average. Indeed it is possible to conceive of cases where the

progress of the class synchronizes with the progress of the dullest pupils. For if a teacher waits until all the pupils have grasped the matter in hand, he must wait for the dull ones. And that means that the brighter pupils, even the average pupils are handicapped by the presence in the same class of those who are slower of comprehension. On the other hand an injustice is usually done to the backward child, for it seldom happens that a teacher holds the class back until the dullest child has understood and learned everything. Fortunately there are few cases in which the dullard is allowed to set the pace for the class as a whole. And just because there are not many instances of this kind, the dull child is often left hopelessly behind. It is possible that there is a good deal yet which he could learn, if he were allowed to proceed more slowly. But being in a class where the average I.Q. is 100 when his own is only 70, let us say, he simply cannot go along at the average rate of the class. If the sub-normal child, however, be placed in a class with other pupils of the same quotient of intelligence, or if he be given individual instruction, he can make a great deal of progress and eventually perhaps become a useful citizen.

The remarkably large percentage of feeble-mindedness that is to be found among the criminal and delinquent classes constitutes a problem of vital public concern. If attention to this problem possesses the possibility of reducing the numbers of these classes by a large percentage, then surely in the interest of public welfare it is the duty of the State to take a practical interest in the problem. It will be of value at this point to refer to the general distribution of school-children in accordance with intelligence quotients, for the majority of criminal and delinquent adults were at one period school-children. The following table is quoted from Woodworth's *Psychology* (p. 274):

	Per cent.
Intelligence quotient below 70,	1
" " 70—79	5
" " 80—89	14
" " 90—99	30
" " 100—109	30
" " 110—119	14
" " 120—129	5
" " over 129	1

In accordance with that table, the general distribution is 60 per cent normal, 20 per cent abnormally bright, and 20 per cent feeble-minded to some degree.

Dr. Leta S. Hollingworth has made an investigation of the classification of feeble-mindedness according to sexes. She finds that in almost all cases statistics from institutions show a greater

percentage of males than females. The United States Government report for 1910 shows 11,015 or 53·8 per cent males as against 9,716 or 46·2 per cent females detained. However such statistics do not prove that feeble-mindedness is more common among men than among women. It is however, as Dr. Hollingworth points out, "an index to the degree to which it is easier for one sex to survive outside of institutions than it is for the other."[1] In New York City the Clearing House for Mental Defectives carried on a research which tended to confirm this conclusion. The research also showed that the males brought to this clinic for diagnosis and commitment were of distinctly higher mental status, age for age, than were the females. The figures proved, for instance, that a girl or woman with a mental age of six years survives outside of institutions about as well as docs a boy or man with a mental age of ten or eleven years."[2] The reason for this is to be found in the fact that men and boys are compelled to follow careers involving competition far more than are girls and women. Moreover studies in the psychological conditions of prostitutes reveals the fact that many girls and women with low mentality have recourse to this low type of life as a means of gaining a living.

One of the great values, then, of the intelligence tests is that they afford a technique through which it is possible to detect feeble-mindedness. If this be done regularly and generally in the public schools, especially where there is a system of compulsory education, it is possible to detect all the sub-normals, and to give them special instruction so as to develop every latent power which they possess. Then it is possible to study the individual cases and find out what are the possible forms of employment for each, where the deficiency is not too great to permit of them remaining as active members of a free community. In other cases, where the defects are more marked, the State can provide institutions in which they can be housed so as to prevent them becoming a menace to the community. In addition to that employment can be given within such institutions in accordance with the intelligence which the individual subjects possess. Many such institutions are alreadv in existence and are doing a magnificent public service.

The most notable revision of the Binet tests, other than the Terman revision, is the Point Scale which was published by Yerkes, Bridges and Hardwick in 1915. As to the type of tests used the Point Scale corresponds closely to the original Binet. It uses the Binet tests and adds a few more of the same general type. It differs from the Binet and the Stanford Revision in the method of arriving at a measurement. It does not employ the

---

[1] Psychology of Sub-normal Children, p. 10.
[2] *Ibid.*, p. 10.

method of grouping in accordance with age, but uses instead a method of scoring responses by means of allotting a certain number of points to each test. It proposes a co-efficient of mental ability which is arrived at by determining the ratio of the score for a child to the average score for a child of that age. A table has, however, been worked out whereby it is possible to ascertain the equivalent mental age for each score in the Point Scale, and may be found in Yoakum and Yerkes book on "Army Mental Tests," [1] pp. 96, 97. After the Terman revision, the Point Scale is more largely used than any other scale in existence.

There have been two notable advances made on the Binet method of measurement. The one is the devising of Performance Tests, and the other of Group Tests. In addition to these there are the tests of achievement which are not so much a difference in method as of application.

Reference has already been made to the fact that one of the chief criticisms of the original Binet scale was that of the language difficulty. This was experienced especially in attempts to test the deaf and the foreign-born, although it was felt also in cases where circumstances had prevented the subjects from obtaining the kind of information required to answer the questions. In all of these cases, it is obvious that a test that depended on language was not really a test of intelligence. It was to remedy this that the Performance test came into being. Healy and Fernald were among the pioneers in this field, proposing a group of performance tests in 1911.[2] These men did not try to group their tests in the form of a scale, but used them for diagnostic purposes as supplementary to the existing scales. H. A. Knox was faced with the problem of the foreign-born in his work at Ellis Island New York, where he had to examine the immigrants many of whom were ignorant of English. To meet that difficulty he devised a number of performance tests which he arranged in the form of a scale. Many of these tests proved to be excellent and have been widely used by other workers. Then came the Scale of Performance Tests from Professors Pintner and Patterson in 1917 with which we shall have occasion to deal more fully later. The aim which these men set before themselves was the selection of a number of tests which called for manipulations involving various capacities and abilities as are included in general intelligence. They felt that "in addition to this principle in the selection of tests, there was the other principle which follows from our general definition of intelligence as the capacity of adjusting to relatively new

---

[1] New York: Henry Holt. 1920.

[2] "Tests for Practical Mental Classification" in Psychological Monographs, Vol. XIII, No 2, Whole No. 54, 1911.

situations, the principle, namely, that each test should present a relatively new situation to the child."[1] The third criterion which they set before themselves follows, of course, from the defect manifest in the Binet tests and its revisions, the language defect. The test must be so arranged that it will be possible for the child to proceed to its solution at a given gesture with no use of language whatever. In regard to the method of grading, these authors sum up the various methods in vogue and state the advantages and disadvantages of each. In conclusion they lend their support to the percentile method, a method which was first introduced by Woolley in 1915. The method is the outcome of the presentation of the results of tests where there has been a large number of persons tested, and it is desirous to know how the group distributes itself. The individual is graded in accordance with his relation to similar performances of others of his own age. Cross comparisons can be made also in respect to the results in the various tests. The authors give tables (pp. 187—198) for the average scores for the various ages from 5 to 14 in the case of each of the 22 tests employed.

In the tests which were employed by the American Army Division of Psychology, the performance tests were found to serve a useful purpose in the testing of foreign-born men who had not yet obtained a good working knowledge of the English language. In that case they were employed if the other tests, involving the use of language were found to be inadequate. The Army Performance Scale is described by Yoakum and Yerkes as " in the main a product of military experience and effort." It consisted of ten tests, which were given numerical scores, afterwards translated into letters in accordance with the army system.

The second of the great advances made upon the original Binet plan for measuring intelligence was the innovation of Group Tests. Where there is a large number of people to be tested, in the interests of the conservation of time and energy it is convenient to be able to test them together. This is possible where the group is composed of people who can read printed directions, providing a careful selection of standardized questions is made. Obviously it is not feasible in the case of foreigners, illiterates, and young children, unless the questions be given orally or by gestures.

The Great War brought about a situation where the usefulness of the group method of testing was apparent on account of the large numbers who had to be examined speedily. It was in the American Army that the first extensive use was made of this method. Soon after the entrance of America as a belligerent, a

[1] A Scale of Performance Tests, p. 21.
[2] Army Mental Tests, p. 18.

meeting of the American Psychological Association was convened which appointed several committees to prepare for action. Simultaneously the National Research Council appointed a committee for Psychology. So at the very outset of American participation the psychologists of the country were prepared for united action. Many of the British and French psychologists served valiantly as individuals, but this opportunity for united action meant an opportunity for more far-reaching service from the psychological point of view. At first a committee of seven experts in mental measurement under the lead of Prof. Robert M. Yerkes was organized to prepare for action. These men worked together for a month, devising ways and means, and at the end of that time, in August 1917, were able to make recommendations to the Surgeon-General in regard to methods for use in the army. "The purposes of psychological testing," as defined in the official medical recommendation, "are (*a*) to aid in segregating the mentally incompetent, (*b*) to classify men according to their mental capacity, (*c*) to assist in selecting competent men for responsible positions."[1] It is informing to read the statement of Prof., then Major, Yerkes as to what was actually accomplished. He lists seven achievements:—

(1) The assignment of an intelligence rating to every soldier on the basis of systematic examination.

(2) The designation and selection of men whose superior intelligence indicate the desirability of advancement or special assignment.

(3) The prompt selection and recommendation for development battalions of men who are so inferior intellectually as to be unsuited for regular military training.

(4) The provision of measurements of mental ability which enable assigning officers to build organizations of uniform mental strength or in accordance with definite specifications concerning intelligence requirements.

(5) The selection of men for various types of military duty or for special assignment, as for example, to military training schools, colleges, or technical schools.

(6) The provision of data for the formation of special training groups within the regiment or battery in order that each man may receive instruction suited to his ability to learn.

(7) The early discovery and recommendation for elimination of men whose intelligence is so inferior that they cannot be used to advantage in any line of military service.[2]

---

[1] Army Mental Tests, p. xi.

[2] *Ibid.* pp. xii. xiii.

I have already indicated that it was in connection with the work of the psychological division of the American Army that the first extensive use of group tests was made. There were two distinct tests used, the one known as *Alpha* and other as *Beta*. The former was devised for men who were fairly literate in the English language; the latter was for those who were not literate in English. The former contained eight tests and the latter seven. But the *Beta* tests were as far as possible "the *Alpha* tests translated into pictorial form so that pantomime and demonstration may be substituted for written and oral directions."[1] Indeed *Beta* was so devised that it could be responded to by men who knew neither how to read nor to understand the English language. Each examination required about fifty minutes to be administered, and the marking was done by a method approximating to the Point Scale. When there were cases about whom there was any doubt after the group test had been given and evaluated, then the examiners were allowed to give individual tests, either the Terman or the Point Scale tests, as they chose. When the armistice was concluded the psychological division had given tests to 1,726,966 men. Such a large volume of data has meant a great deal for the science in enabling us to further standardize tests and to reach conclusions regarding the results. There are a good many parallels to be drawn between an army training camp and a school, and we shall have occasion again to revert to this work.

A further extension in the work of measurement of mental abilities is to be seen in the standardization of measurements of progress and of special abilities. The need for this type of technique has developed from the experience of uneven standards in examination. An interesting account is given by Monroe[2] of an investigation made by Starch and Elliott into the accuracy with which teachers mark papers in geometry. "A facsimile reproduction was made of an actual examination paper in plane geometry. A copy of this reproduction was sent to each of the high schools in the North Central Association of Colleges and Secondary Schools, with the request that it be marked on the scale of one hundred per cent by the teacher of geometry. The teacher was asked to mark the paper by the methods he was accustomed to use. Papers were returned from 116 schools and the results tabulated. When we consider that the subject-matter of geometry is quite definite, and that the papers were marked by teachers who were thoroughly acquainted with the subject, it would seem that we might expect the mark or grades placed upon the examination paper to be in close agreement. However, exactly the opposite was the case . . . Of the 116 marks, two were above 90, while

[1] Mental Tests, pp. 16, 17.

[2] Measuring the Results of Teaching, pp. 8 and 9.

one was below 30. Twenty were 80 or above, while 20 others were below 60. Forty-nine teachers assigned a mark passing or above, while sixty-nine teachers thought the paper not worthy of a passing mark." This type of evidence was repeated by the same investigators in the cases of other subjects, and by many others who have carried on similar investigations. The result of this conviction in regard to the inaccuracy of school marks has been a growing effort in the direction of standardizing examinations. This type of test varies somewhat from the others in that the tests which we have been discussing are intended as measurements of the subjects' intelligence, while these are calculated to measure the results of teaching or the subjects' achievements in special lines. As a result of work in this branch of the subject we have now a number of tests in operation for the measurement of arithmetical ability, ability in spelling, in reading, in geography, history, foreign languages, etc. The work of standardization, as in the case with the Binet tests, is done on the basis of experiments upon thousands of pupils and the tabulation of results.

To be sure Binet made a beginning in this type of test also in his *barème d'instruction.* With his death the work seems to have come to a standstill in France. But it has been carried on with a good deal of vigour in America by such men as Thorndike, Judd, Monroe, Starch, Elliott, Ayres, Courtis and others. In England it has been taken up by such investigators as Ballard and Burt. The American tests are arranged on scales corresponding to grades, whereas the Englishmen favour the age scales. But that makes standardization more difficult, as it creates a situation parallel to the use of the metric system in France and the old linear measurements in England, or to the sterling currency in Britain and the decimal currency in America. It is to be hoped that a similar breach will not be permitted to persist in educational measurements, but that the workers in the field will come to an agreement as to the adoption of a common scale.

CHAPTER II.

THE OBJECTIVE IN MENTAL MEASUREMENT.

One of the most fundamental questions which confronts us as soon as we take up the discussion of psychological tests of mental abilities is, What is it that we are trying to measure? We are quite accustomed to the idea of measuring cloth or land or temperature. But the application of the technique of measurement to psychological matters is something new. And it demands some careful deliberation. It touches the problem of the legitimacy of such a process. And further it involves careful consideration on account of the complex nature of many of our mental processes and abilities.

At the same time we need not have a precise definition [1] of the nature of intelligence before we begin our process. Indeed there are some who doubt the probability of ever achieving a satisfactory definition. Prof. L. P. Jacks, says: "I doubt if we shall ever be able to produce an intelligent definition of intelligence."[2] There are some who are so obsessed with apriorism that they resent the idea of undertaking any kind of experimentation unless they have a clear conception of that upon which they are going to experiment. Stern very appropriately reminds us that this type of objection is irrelevant, for it is not the method of science, and in this task we must proceed by the best approved methods of science. He reminds us therefore that "We measure electro-motive force without knowing what electricity is, and we diagnose with very delicate test methods many diseases the real nature of which we know as yet very little." On the analogy of other scientific investigations he therefore argues quite relevantly that "progress in testing intelligence may shed light from a new angle upon the theoretical study of intelligence and thus supplement the psychology of thinking in a valuable manner. If it turns out, for instance, that certain symptoms are relevant and others irrelevant for the differentiation of the intelligence shown by different persons; if, again, one series of these symptoms exhibit a high degree, another series a low degree of intercorrelation, then our knowledge of the structure of intelligence must thereby be little by little increased, and thus there will develop a fruitful reciprocity between the two phases of investigation, theoretical and applied."[3]

---

[1] The reader is referred to a symposium on what is meant by "Intelligence" which appeared in several issues of *The Journal of Educational Psychology* in 1921. Chapter XVI in Dr. P. B. Ballard's *Group Tests of Intelligence*, is also a useful discussion. See also C. Spearman, *The Nature of "Intelligence" and the Principles of Cognition.*

[2] From the Human End, p. 55.

[3] The Psychological Methods of Testing Intelligence, p. 2.

At the same time, as Stern acknowledges, it is not possible to begin an investigation of this nature without some previous conception of the nature of that which we are investigating. So long as we regard the definition with which we begin our work as a hypothesis, possible of modification in the light of the facts that will be brought to light, we shall be guarding ourselves against the dangers of deduction. In other words, we must follow here the trial-and-error method of the scientific laboratory, for, to be sure, ours is a laboratory though it takes the form of a school-room or of an institution for defectives. At the same time there has been so much work done that we are by no means in the dark as to the nature of intelligence. As a result of the immense amount of work that has been done both in theoretical and experimental psychology, we are able to begin with a definition or an analysis that is fairly well attested. It is even possible that our investigations will serve rather to confirm than to compel us to modify our hypothesis.

We shall begin by a consideration of those elements which enter into intelligent behaviour, and then later consider the problem of definition. In the first place it is to be observed that there is no such thing as intelligence. To use the word in the sense of a thing or an entity is a mistake. It is more to be used in the descriptive sense as applicable to certain actions, behaviour, tendencies, dispositions, rather than in the substantive sense as a faculty or department of the mental life. Intelligent reactions are to be differentiated from the reflexive and instinctive types by the presence of conscious adjustment which the other two do not involve. Intelligent reaction involves the functioning of the cerebral cortex whereas the other types involve only the lower brain-centres. Our interest then is not in the delineation of the qualities inhering in a substantial intelligence, but in the discovery, as far as we can, of the characteristics of those reactions which we describe as intelligent. It must be with these limitations implied, if not repeatedly expressed, that we make use of the word *intelligence.*

This is by no means a merely theoretical problem for the psychologist, but is of practical importance to us in connection with the analysis of tests. If we are clear in our thinking in regard to the elementary factors of intelligent conduct, then we can study to devise tests that will examine the various factors, and in a complete test may guard against the possibility of some important factor being left untested. Intelligence is much too complex for us to expect ever to devise a single test that will measure it or gauge it. But by means of a variety of tests we are able to examine the various factors, and thus measure the totality by means of the parts. In that way, it is important to attend to the particular function of each separate test.

At the same time it will be observed that we are here examining psychological tests of mental abilities—a phrase of wider connotation than intelligence. For the application of the mathematical method has not been confined, as we saw in the first chapter, to intelligence. It has also been applied to the measurement of attainment through the standardization of examinations. A clear discrimination of the purpose of the test being employed is one of the first prerequisites for its intelligent use. It will be the part of wisdom for those who take up this work in a practical way to form the habit of asking themselves: What am I trying to measure?

In the main, there are two different points of view in regard to the nature of intelligence. The one is reflected in Binet, Spearman and others, and maintains that there is such a mental phenomenon as general intelligence. The other theory which is defended by Thorndike and his school is that there is no general intelligence but that there are particular intelligences, or better, mental abilities, which are independent of one another. Both schools have reached their conclusion from the same data, the divergence being one in interpretation. These points of view have an important bearing on the method and character of the tests.

I have said that Binet held to the doctrine of a general intelligence. He said: "It seems to us that in intelligence there is a fundamental faculty, the alteration or the lack of which is of the utmost importance for practical life. This faculty is judgement, otherwise called good sense, initiative, the faculty of adapting oneself to circumstances. To judge well, to comprehend well, to reason well, these are the essential activities of intelligence."[1] Again in dealing with *L'Intelligence des Imbeciles* in *L'Année Psychologique*, 1909, the same authors entered into a more elaborate discussion of the nature of the higher mental processes. In justice to Binet it ought to be said that the later article shows the evidence of more mature psychological judgement, and smacks much less of the old faculty method. It was one of the prime merits of Binet that he was ready to move from his positions whenever he realized that his investigations had brought to light data which made his earlier positions untenable. In his later work, Binet describes the features of intelligence as (1) the tendency to take and to maintain a definite end or direction; (2) the capacity to make adaptations in pursuance of the directing end to be attained, which guides the subject even unconsciously; and (3) the power of auto-criticism whereby the person can judge of what has been done with reference to the end and to the standard. These three aspects of intelligence are shown as operative in the performance involved in such a test as the re-arrangement of the

[1] Binet and Simon: The Development of Intelligence in Children, translation by Kite, p. 42.

disarranged parts of a rectangle, known as "the patience-test." Here (1) the end or direction is the figure that is to be re-formed, (2) the adaptation is in the trials of various combinations in the process of striving towards the end, and (3) auto-criticism comes out in the judgements made on the trials made with reference to the model, so as to determine which is correct. An examination of the Binet tests will show that many of them are devised so as to test these three factors, as e.g., the paper-cutting test, the re-arrangement of dissected sentences, the copying of drawings from memory, the indication of omissions from pictures, etc.

Spearman and Hart agree that there is a mental activity which may be designated as intelligence. They regard general intelligence as a "common central factor" or "central tendency," not lending itself to exact definition, but which participates in a greater or less degree in special mental activities, indeed in mental activities of all sorts.[1] Spearman made a study of several special abilities such as adding numbers, memorizing words, and others, and compared the results after trying the tests on a number of subjects, correlating the results with one another. It was observed that as a rule a subject which was good in one thing was also good in other things. Not many people are able in one direction only. Moreover he observed a fairly high degree of correlation between the various abilities in the subjects tested. That led him to the conclusion that there is a sort of general store-house of intellectual power from which the person is able to draw for the particular needs, a general intelligence, or general ability.

The German psychologist E. Meumann was also an adherent of the doctrine of general intelligence. He offered a dual definition psychological and practical. From the psychological view-point, it is the "capacity for independent, productive thought" whereby new mental products may be created out of the data supplied by the senses and memory. From the practical point of view, it is "the intensity of the whole mental life" which functions in the correction of mistakes, the overcoming of difficulties, and in adaptations to environmental conditions.[2]

Ebbinghaus in his *Grundzuge der Psychologie* holds to the same general theory. He says: "Intellectual ability consists in the elaboration of a whole into its worth and meaning by means of many-sided combinations, correction and completion of numerous kindred associations . . . It is a combination activity."[3] He regarded intelligence in that way as a unifying comprehending function whereby heterogeneous parts which are in themselves

---

[1] General Ability ; Its Existence and Nature, in the British Journal of Psychology, Vol. V, 1912-13, pp. 51-84.

[2] Experimentelle Pädagogik, Vol. II, pp. 102 ff., Leipzig, 1913.

[3] Vol. II, Leipzig, 1913.

largely disparate are regarded homogeneously. It is a function which includes the abilities to abstract, compare, contrast, and classify.

Mr. Cyril Burt gives it as the conclusion of his investigations that there is a strong suggestion "that it is one feature or function of attentive consciousness which forms the basis of intelligence, namely, the power of readjustment to relatively novel situations by organizing new pscyho-physical co-ordinations."[1] Mr. Burt was one of the earliest and one of the most persistent investigators in the field, and he concluded that almost any kind of ability correlated fairly closely with intelligence, but that the correlation was much closer in some instances than in others. "Of all the tests proposed," Ballard quotes him as saying, "those involving higher mental processes, such as reasoning, vary most closely with intelligence."[2]

Ballard lends his support to the theory of a general factor of intelligence. He is impressed by the reasoning of Spearman in regard to the correlation between various mental abilities. To quote his own words: "Generally speaking, a wise man is wise in all things, a fool is a fool all round. Indeed, it can be proved mathematically that there is a positive correlation between all forms of native ability; they always tend to hang together; the odds are always in favour of high ability in any given function being accompanied by high ability in any other function. Why should this be? Why should mathematical ability be correlated, as it is, with linguistic ability? Even if we make every allowance for such operations as might be common to two abilities, we still fail to account for the whole relationship. There still remains an unexplained nexus. We are forced, in fact, to assume a general factor common to all the multifarious operations of the mind, a factor with which each special ability is, in its own measure, charged and energized. This common factor is intelligence."[3]

One other writer may be referred to as holding to the theory of general intelligence, viz., W. Stern. I have already alluded to the fact which Stern recognized, namely, that any definition which is made at the outset of an investigation must be in the nature of a working hypothesis, rather than a categorical apriorism. With that qualification to safeguard his position, Stern then gives his definition of intelligence as "a general capacity of an individual consciously to adjust his thinking to new requirements: it is general mental adaptability to new problems and conditions of life."[4] The author then proceeds to a defence of the terminology which he

---

[1] Experimental Tests of General Intelligence, in the British Journal of Psychology, Vol. III, 1909–1910, pp. 94—177.

[2] Mental Tests, p. 27.

[3] *Ibid.*, p. 25.

[4] Psychological Methods of Testing Intelligence, Whipple's Translation, p. 3.

has employed, claiming that by it he has successfully differentiated intelligence from other mental abilities. With reference to the conception of intelligence as a general mental ability he says: "The fact that the capacity is a *general capacity* distinguishes intelligence from *talent* the characteristic of which is precisely the limitation of efficiency to one kind of content. He is intelligent, on the contrary, who is able easily to effect mental adaptation to new requirements under the most varied conditions and in the most varied fields. If talent be a material efficiency, intelligence is a formal efficiency."[1] Again towards the close of his monograph the author in his advice to teachers suggests that they bear in mind the conception of intelligence as a "general mental adaptability to new problems and conditions of life." In so doing he advises them to attend particularly to the word "general," and to "guard against identifying with intelligence any sort of special ability or the mere possession of information or readiness in speech. Because of the general nature of intelligence it is essential to take into consideration the way in which the child behaves in quite different situations and when confronted by problems of various sorts."[2]

Professor E. L. Thorndike disagrees with those who hold to the doctrine of a general intelligence. His method of research was, like that of Spearman and Hart, mathematical. He investigated specific mental abilities such as the addition of numbers, the discrimination of lengths, the memorization of words, and the sorting of cards. Then he compared the results, noting the facts in regard to correlation between the various abilities and the degrees of variability. His conclusion was the exact opposite of Spearman from the same type of investigation. There is no such things as general intelligence; all that we can observe are particular intelligences, individual abilities. Thorndike found that the correlation between particular abilities showed very poor correlation. One student may be a good linguist and hopelessly poor at mathematics. Another may be brilliant in poetry and stupid in exact science. Thorndike made an investigation of the comparative mentality of dependent children who are the inmates of charitable asylums with ordinary public school children. The tests were of two kinds, the one involving language and the other calling for mechanical ingenuity. It was found that the disparity between the two groups was much more apparent in the tests involving language than in the performance tests. To be sure there are two ways of interpreting that result: the one is to say that the performance test is a much more reliable test of intelligence than the test which requires the use of language; the other is to conclude,

[1] Op. cit., p. 4.
[2] *Ibid.*, p. 120.

with Thorndike, that abilities are specific, and that an individual or a group may do much better in one test than another because they have a higher type of ability in the one direction than in the other.

Dr. Hart and Professor Spearman claimed that the result of the various tests was the disclosure of a perfect hierarchical order among the correlation co-efficients. They employed the mathematical method in their investigation. It is unnecessary for our present purpose to go into the details of the mathematical formulas and their workings. Those who are interested may find a full discussion in Brown and Thomson's *Essentials of Mental Measurement*, chapters 9 and 10. The criticism of those authors is that Spearman has used a simplified formula in arriving at his correlation which yields a result supporting his theory, a formula which these scholars deem does not do logical justice to the data. Thorndike and Brown carried on independent investigations following the publication by Spearman of his findings. They found results which conflicted very radically with those of Spearman. Thorndike made his calculations on the basis of tests of accuracy in drawing lines, equal to given lines, in filling boxes with shot equal in weight to standard weights, and on judgements of general intelligence made by fellow-students and by teachers. He found that there was a much higher correlation between the discrimination of weights and the discrimination of lengths than there was between either of them and general intelligence. The co-efficients were:

Accuracy in the discrimination of lengths and intelligence	0·15
Accuracy in the discrimination of weights and intelligence	0·25
Accuracy in the discrimination of weights and of lengths	0·50

Thorndike's comment on the results of the investigation was as follows: "In general there is evidence of a complex set of bonds between the psychological equivalents of both what we call the formal side of thought and what we call its content, so that one is almost tempted to replace Spearman's statement by the equally extravagant one that there is *nothing whatever* common to all mental functions, or to any part of them."[1]

Professor Spearman continued the investigation in collaboration with Dr. Hart, and, while recognizing the difficulty of the investigation, returned to the original conclusion that there must be a general factor which we call intelligence to account for the

[1] Thorndike, Lay and Dean: The Relation of Accuracy in Sensory Discrimination to General Intelligence, in American Journal of Psychology, July 1909, Vol. XX, p. 368.

perfection in the coefficients of correlation. The conclusion was based on their observation and calculation of a hierarchical order among the coefficients of correlation. They believed that without such a general factor, the average correlation between the various abilities would be either zero or negative. Brown and Thomson bring forward some rather damaging evidence to this position, by showing that it is possible to produce a hierarchical order by the random overlap of group factors without any general factor present. Their experiment is that of drawing a card from a pack of playing cards, replacing, and shuffling before each draw, and then proceeding to identify the group factors of each variate by using a single suit of the pack. "From these, and from the total number of factors both specific and group in each variate, can be found the correlation which would occur between the variates were we to throw dice, one to each factor, and repeat the throwings a large number of times."[1] The experimenters carried out such an experiment and worked out the results which showed evidence of a remarkably high degree of hierarchical order. In accordance with the criterion of Spearman and Hart there must then be present a general factor, whereas the facts are that the whole procedure was random.

Thomson and Brown, having disposed of the Spearman theory of a general factor, on the grounds of incorrect mathematics and of having set up arbitrary standards, proposed instead "a sampling theory of ability." They prefer to think "of a number of factors at play in the carrying out of a mental test, these factors being a sample of all those which the individual has at his command."[2] This theory "does not deny General Ability, for if the samples are large, there will of course be factors common to all activities. On the other hand it does not assert General Ability, for the samples may not be so large as this, and no single factor may occur in every activity. If, moreover, a number of factors do run through the whole gamut of activities, forming a general factor, this group need not be the same in every individual. In other words General Ability, if possessed by one individual, need not be psychologically of the same nature as General Ability possessed by another individual. Everyone has probably known men who were good all round, but Jones may be a good all round man for different reasons from those which make Smith a good all round man. The Sampling Theory, then, neither denies nor asserts General Ability, though it says it is unproven. Nor does it deny Special Factors. On the other hand it does deny the absence of Group Factors."[3] In defence of the theory the authors

---

[1] Brown and Thomson: Essentials of Mental Measurement, p. 176.

[2] *Ibid.*, p. 188.

[3] *Ibid.*, p. 189.

point out that it is in agreement with the line of thought which has proved fruitful in other sciences. "Any individual is, on the Mendelian theory, a sample of unit qualities derived from his parents, and of these a further sample is apparent and explicit in the individual, the balance being dormant, but capable of contributing to the sample which is to form the child." [1]

There is thus a lack of unanimity in regard to the formal question. Yet the accumulation of evidence seems to give weight to the theory that there is no such thing as a general power of intelligence which can be directed at pleasure, now to one object and now to another. It is not safe to conclude that exceptional ability in one direction will be accompanied by special ability in another or in all others, or even that improvement of one ability will carry a corresponding improvement in other abilities. If there be any such corresponding improvement it will be due to the two abilities making use of common forms of perception, attention, and so on.

Our English word *intelligence*, like the word *intellect*, is a derivative of the Latin *intelligere*, to understand. (So also the French, *intellect* and *intelligence*, and the Italian *intelletto* and *intelligenza*, and the German *intelligenz*). The sense in which the word is used in the earlier psychologies was that of the cognitive faculty. The tendency grew to use the word *intellect* rather in regard to the distinctly conceptual processes, and in that way a distinction arose between the words intellect and intelligence. This distinction is being found to serve a very useful purpose in Comparative Psychology. As Stout and Baldwin have pointed out: "We speak freely of 'animal intelligence'; but the phrase 'animal intellect' is unusual." [2] Lloyd Morgan accepts the distinction, and elaborates it by saying that "the term (intelligence) may be conveniently restricted to the capacity of guiding behaviour through perceptual process, reserving the terms intellect and reason for the so-called faculties which involve conceptual process." He however makes this reservation that "it is probably best for strictly psychological purposes to define somewhat strictly perceptual and conceptual (or ideational) process and to leave to intelligence the comparative freedom of a word to be used in general literature and therein defined by its context." [3] Lloyd Morgan then proceeds to show that comparative studies have brought to light the deficiency of animals when it comes to such analysis and abstraction, even in simpler forms, which are required for conceptual thinking. Animals are however capable of perceptual intelligence. Associative representation enables them to

---

[1] Op. cit., p. 190.

[2] Art. *Intellect* or *Intelligence* in Dictionary of Philosophy and Psychology.

[3] Art. *Intelligence* in the Encyclopædia Britannica, 11th edit., Vol. XIV, p. 681.

learn. Experiments have been performed in the comparison of the learning abilities of men and lower animals and the results are not always very complementary to the human animal. But as soon as tests are applied which demand abstraction, be it never so simple, the lower animal is at once at a disadvantage. For that reason the measurement of intelligence in the lower animal cannot be attained by the type of tests which we are now considering.

We may now move to a consideration of some of the factors which are samples of what the individual has at his command, as Thomson puts it, some of the factors of intelligence. Such an analysis is as possible as it is because so much work has been done in the testing and measuring of mental abilities.

In the first place intelligence involves, as Woodworth points out, doing a miscellaneous lot of things and doing them right. Both Spearman and Thorndike had observed that fact, the difference being in the interpretation which they gave to it, the former holding that there was a general factor which determined one's ability or otherwise for doing them, while the latter said that each distinct thing demanded its own specific ability. How this complex nature is to be conceived is thus no simple question. There is however fair unanimity in regard to the main fact, though we may have to admit that the power of intelligence attains only an approximate measure of uniformity. Even those who hold to the theory of a general intelligence have to admit variations among persons, and variations in the abilities of the same individual, some being scored higher than others. At the same time the fact of the unity of the mental life makes it apparent that it is not possible to set any one mental ability apart from all of the others and to measure it. The interpenetration of the various parts of life make it unavoidable that we should measure other elements when we set out to measure any one. That does not militate however against the possibility of devising tests which shall have in view the testing of certain functions or abilities, even if other factors are brought into play at the same time.

The complex nature of intelligence may be illustrated by reference to the literature of experimental psychology. The experimental psychologist works on a method somewhat different from the test method which we are considering. The difference has to do largely with the objective, in the case of experimental psychology the aim being more theoretical, and in the case of educational psychology more practical. The former wants to make such careful observations as will assist in the formulation of hypotheses or principles, while the latter seeks to diagnose mental illnesses, and to afford a criterion whereby subjects can be classified for practical considerations such as the organization of a school or the protection of a community from the harms resulting from lack of control of feeble-minded people. The experimentalist

takes into consideration introspective factors, whereas the educationalist deals only with tests of overt behaviour. At the same time the experimental results are not without their significance for the educationalist, as furnishing data concerning the processes which he is studying. Psychological research brings to light certain specific facts of which the educationalist may well take cognizance in analyzing his results. For example the intelligence test in certain cases discloses a mental condition which is abnormal, the mental processes appearing to be slow and sluggish though not quite stupid. The result seems to warrant a conclusion that the mental abnormality is symptomatic rather than a congenital deficiency. The investigator knows that the effects of certain drugs or of certain poisonous gases are likely to produce symptoms such as appear in the subject, and investigates the environment from which the subject comes as well as his habits, and has at once a clue to a correct diagnosis. In this way it will appear that the broader the knowledge of the investigator concerning theoretical and experimental psychology, the safer he will be as a conductor of mental tests. The processes are too intricate and some of them too complex, and the life of the child much too important for it to be safe to allow any individual who happens to have read a book or two on the subject and learned the method of scoring to conduct a test on which the future of the subjects is to depend. Much as we desire to see this work undertaken in real earnest here in India, we cannot too strongly warn one another against the dangers which are involved in inviting indiscriminate testing on the part of untrained enthusiasts.

As a second characteristic of intelligent behaviour, I would point out that it is always purposive. It is conduct with reference to some end of which the individual is conscious. It was the merit of the late William James to have pointed that out long before psychological tests of intelligence had been dreamed of. In his *Principles of Psychology* he says "The pursuance of future ends and the choice of means for their attainment are the mark and criterion of the presence of mentality in a phenomenon. We all use this test to discriminate between an intelligent and a mechanical performance. We impute no mentality to sticks and stones because they never seem to move *for the sake of* anything but always when pushed, and then indifferently and with no sign of choice." He then alludes cogently to that problem of philosophy, as to whether or not the cosmos is an expression of an intelligent power or the outcome of blind mechanical laws of necessity. "If we find ourselves, in contemplating it, unable to banish the impression that it is the realm of final purposes, that it exists for the sake of something, we place intelligence at the heart of it and have a religion." [1]

[1] Vol. I, p. 8.

By purposefulness we mean the ability consciously to adapt to ends. We have referred to the fact that Binet took this to be the principle characteristic of intelligence. In his three-fold division of the nature of intelligence, he has first the consciousness of the end to be attained, second the trial of possible means to that end, and thirdly auto-criticism of the trials made. It is true that his tests actually tested a much wider range of abilities than those which he so mentioned, but it is significant that he considered this ability of adaptation as the outstanding element in intelligent processes.

There are some tests which are particularly well suited to test one's power of adaptability. One which Binet refers to as well arranged to test this power is the patience puzzle. Two rectangular cards of the same dimensions are taken, one of which is cut into two triangular pieces by cutting along one of the diagonals The uncut card is placed on the table, one of the longer sides towards the child, and by its side the two triangular pieces. Then the examiner tells the child that he is wanted to take the two triangular pieces and put them so together as to look like the uncut card. This test has been tried at Saidapet with children of six to seven years, and each child given three trials of one minute. Miss Gordon reports that "the bright child sometimes fails but usually not without many trials combinations which he rejects as unsatisfactory. The dull child often stops after he has brought the pieces together into any sort of juxtaposition, however absurd and may be quite satisfied with his foolish effort. His mind is not fruitful and he lacks the power of auto-criticism." [1]

There are other simple tests that are well adapted to test adaptability. Take for example the drawing of a square or of a diamond from memory. Obviously the end is the production of a drawing resembling the copy which the child is shown. The effort to produce a drawing which will bear resemblance to the original is an attempt at adaptation, and attempts at correction or improvement are evidences of self-criticism.

Many of the more complex processes of life involve the calling into play of this power of adaptation. A number of the individual tests are illustrative. Take for example the substitution test which Whipple describes.[2] This test is administered in various forms, the main point of which is that the subject is asked to substitute one set of characters (letters, digits, familiar geometrical forms, etc.), for another set of characters in accordance with a plan set before the subject in printed directions. The principle admits of

---

[1] Teachers' College, Saidapet, Bulletin No. 15, pp. 16, 17.

[2] Manual, Vol. II, pp. 133–150.

many variations, but the nature of the test and the presence of directions involves the necessity of the subject holding before consciousness the end in view and deliberately setting about the adaptation of means to that end.

Purposiveness implies the comprehension of meanings. Certain tests have been devised that are especially useful in examining the subjects' ability to rearrange unrelated fragments into meaningful forms. Such a test is the completion test devised by Ebbinghaus, and afterwards modified in various ways. The subject is given a paragraph from which certain words are omitted and is asked to complete the paragraph by filling in words which will make sense. Ebbinghaus omitted syllables in some cases, but Terman thought it better to omit whole words as it would not then depend on the child's ability in word-analysis. Ebbinghaus' defence of this test is that it brings into play the essential factor of intelligence, namely combinative activity. Ability to combine into a significant whole parts, which independently seem to be unrelated or even to give the impression of contradiction, involves that creative ability of combination which is the essence of intelligence. Whipple shows that in the case of the completion test, as also in the case of the substitution test, there is a high degree of positive correlation with intelligence. To be sure it would have the same defects which characterise any language test, but allowing for limitations of that kind it is a well verified test.

Another type of test which is well suited to examine one's power of adaptation is the form-board test. The device is in the form of a board with holes cut in in the form of geometrical figures. There are various shaped blocks which if put together in the correct way may be fitted into the geometrically formed holes. This test has been found by many investigators to be exceedingly useful as a test of native ability, especially as diagnostic of the child's ability to deal quickly and well with a new situation. Several investigators have found it a very quick and accurate means of differentiating the normal from the feeble-minded. Attempts to place the blocks in holes where it is manifestly impossible for them to go, and then turning them up-side-down or otherwise trying to manipulate them so that they will go where they cannot go has been found to be symptomatic of defectiveness. The ability to perceive form and the rapidity with which the movements are executed are good indications of the degree of mentality. The Rev. D. S. Herrick of Bangalore has carried on some investigations with the Goddard form-board, and believes that it is a very good type of test with which to begin work here in South India. Certainly it is better until the language difficulty is obviated by translations and adaptations of the language tests being made available in the vernaculars.

Adaptation means responsiveness to relationships. The type of test that is to measure the ability of a person to respond to a stimulus is the type that will compel the subject to face a new situation. Otherwise the test might be more one of memory than of intelligence. Stern and others lay a great deal of stress upon that factor. By definition, Stern has indicated his conception of intelligence as involving one's capacity for adjusting oneself to new requirements, new problems, and new conditions. It has to do, thus, with a person's external relations, and the manner in which he is able to adjust himself, his thinking, and his conduct to new requirements. Mr. Herrick found that the form-board served this purpose remarkably well. He examined more than 700 children and says: "Not one of the more than 700 boys and girls tested had ever seen a form-board, it is safe to assert. Few, if any, of them in all probability had ever handled blocks of wood or other material of different shapes, much less tried to fit them into holes of corresponding shapes. To be confronted with a block full of holes and a lot of blocks, and to be told to put the blocks into the holes as quickly as possible, was a new situation for each of these children. Thus it was well adapted to test their intelligence. At the same time, there was nothing unreasonable in the test, so perfectly simple is it."[1]

A third element in intelligence is the presence of the voluntary phase. It is at this point, of course, that intelligence is frequently differentiated from instinct. Reflexive and instinctive behaviour are involuntarily performed, whereas intelligent behaviour brings into play conscious conation. That is one reason that a mental test is a real criterion of mentality. If there were no intelligence, and the subject acted only from instinctive tendencies, it would mean that he would not learn, and there might be repetitions of instinctive responses useless if not accompanied with harmful consequences. Let us take the example of a child burning his finger by contact with fire. Instinctively, he withdraws the finger on experiencing the feeling of pain. Not only so; he learns also to associate painful feeling with that type of experience, and so learns to avoid repetitions of that act. Were he equipped with a mechanism for instinctive reactions only, he would doubtless withdraw the hand every time it went into the fire from the instinctive tendency of self-preservation, but he might not learn to avoid repetitions of the painful experience, and at times that might lead to disastrous consequences. When, therefore, a mental test discloses the fact that the person is capable of instinctive reactions only, but not of intelligent responses, we know that there is danger ahead of that person unless he is cared for by the State or some other control.

---

[1] Article in the *Journal of Applied Psychology*, September 1921, reprinted in *Methodist Education*, April, 1922.

One of the best evidences of voluntary power is in connection with attentiveness. Attention is a fundamental form of conation and attention is necessary for conscious control. The "span of attention" correlates very closely with mental ability in general. Attention is a necessary factor for the successful performance of any test. Some of them in particular are an index to one's power of attentiveness. Take such a test as the drawing of a design from memory after the subject has been shown the design only a few seconds. Two such tests are given as ten-year-old tests in the Binet scale, and the children are allowed to look at the design for ten seconds, after which they are asked to reproduce them from memory. Binet points out what must be obvious, namely, that one of the factors demanded for success in this test is attention.

A fourth evidence of intelligence is the tendency to explore. This is the conscious factor which has evolved from the instinctive tendency to pry into the strange and the unusual. Little children, not to mention monkeys, dogs and other animals, show very decided tendencies to explore the unknown. This is often spoken of as the instinct of curiosity. The conclusion of the biologists is that it is, without doubt, one of the primary impulses. It is this instinctive tendency that lies at the basis of the search for knowledge, and scientific research. To it we owe, as Shand says, "most of the disinterested labours of the highest types of intellect. It may be regarded as one of the principal roots of both science and religion."[1] The greater tendency there is to explore, the greater will be the intellectual vigour of the child. It is one of the obvious sources of the craving to increase the stock of one's knowledge by investigation and experimentation. It must be quite clear that success in dealing with many of the tests depends upon the strength of this tendency. If the child is content to make one or two trials and then give up as failed, or if he is content with a half success, he will not do nearly so well with many of the tests as the child who is more of an exploring turn of mind. The form-boards tests give abundant illustration of that. Some children will give up after one or two failures; others will content themselves with getting a block into a hole regardless of whether or not it is meant for that hole and fits it; others will continue exploring with the block in the various holes until they succeed; still others will make their explorations mentally, and after mentally working over the situation will try out their conclusion, and usually with success. The completion test of Ebbinghaus is another test the success of which depends upon the working of this tendency. Inevitably it is present to some degree in all mental operations which are called forth by the tests, and the more developed it is, the higher will be the score of intelligence.

---

[1] The Foundations of Character, p. 59.

Perhaps we may take as a part of the same explorative tendency the factor of persistency. Some psychologists speak of it as the instinct to self-assertiveness. Whether or not it be so classified is not important for us, as the theoretical question is secondary. But we do know that the ability to persevere and to assert oneself as dominant over circumstances and problems is an important element in the attainment of success in mental tests. And it is also of importance in the realization of the self into which all intelligent behaviour is integrated.

A fifth factor which we may note in intelligence is retentiveness. This has been found to have a very high correlation with intelligence. The digit repeating test, e.g., has proved to be most useful in the arrangement of a scale of intelligence, because it is so easily measureable, and because it is an ability that develops correspondingly to general mental development. In the Binet scheme the number of digits repeated is found to correspond to the mental age on the following basis :

A child of three years repeats 2 digits.
A child of four years repeats 3 digits.
A child of eight years repeats 5 digits.
A child of fifteen years repeats 7 digits.

Similar experiments are the repetition of sentences in which the number of syllables increases in proportion to the increasing mental age.

A child of three years can repeat a sentence of 6 syllables.
A child of five years can repeat a sentence of 10 syllables.
A child of fifteen years can repeat a sentence of 26 syllables.

The test of association, both controlled and free, brings into play the ability to retain. The greater the power of retentiveness, the better will be the response to association tests, because there will be on hand a larger stock of associations from which the subject can draw. A person who is rich in associations is found to be a person of high mental ability, whereas the feeble-minded child invariably indicates his abnormality in the poverty of his associations. The correlation is no doubt due to the higher degree of retentiveness which characterizes the more intelligent person. The analogy test is, of course, a special case of association, and, like the other association tests, works well as a test of retentiveness and consequently of intelligence.

Woodworth points out that intelligence includes an element of submissiveness, which we may take as a sixth factor. Perhaps it is a part of the process of adaptation, and need not be considered as a distinct factor. This involves the social factor, and is one of the most difficult elements to be measured. Comparative data have been assembled by Binet, Stern, Terman and others in regard to the intelligence of children from differing social environments, and it has been found that those from the higher social groups test

higher than the others. Terman's investigation shows that the average Intelligence Quotient for the children from a superior social group was 107 whereas that from the lower social group was 93. Some may think that such a result was the outcome of circumstances, and would not persist if the circumstances were improved. Repeated tests seem to indicate that schooling rather accentuates than diminishes the disparity. Whether it would disappear if the children were taken out of the community in which they have been reared and placed in an environment of the higher level is an experiment that has not been seriously tried. Tests made in well-conducted orphanages after children had enjoyed the privileges of the new and improved environment for two years indicate that the difference still remains.[1] There is one point to be observed, namely, that the children from the superior classes are more in the company of adults than is the case with the inferior classes, and that means the greater call for submissiveness. Submissiveness in the sense in which we are considering it is the willingness to yield to the control of others whose superior authority is recognized rather than a yielding to superior force. In that sense it is a mark of self-discipline and control, which is a characteristic more highly developed in the intelligent than in the defective.

It need scarcely be added that intelligence is concerned more with one's native equipment than with something acquired. It is, as Ballard says, "ability as distinct from knowledge, capacity as distinct from content, power as distinct from product."[2] In our measurements we are trying to find out what the mind is capable of doing rather than what it has done. At the same time it would be impossible to measure a contentless capacity. There is no way of measuring the intelligence of the new-born infant. Until there is some knowledge we have no criterion for measuring the ability to attain knowledge. That is one of the difficulties against which the intelligence test has to struggle. If it is to be purely an intelligence test, we do not want to make it a measurement of knowledge, yet we cannot measure intelligence without taking knowledge into account.

The use of standardized tests is not confined to the measurement of intelligence. They are employed as vocational tests. To be sure vocational selections depend largely on the measurement of intelligence. It has been ascertained that there are certain forms of useful, manual employment which are open to imbeciles. Even domesticated animals may be called into service, as we know. There are other useful occupations which demand some degree of intelligence and yet no specialized type, occupations which can be manned by people who are slightly sub-normal and yet by no means

[1] See Terman: The Measurement of Intelligence, pp. 114 ff.

[2] Mental Tests, p. 23.

feeble-minded. There are other types of employment which are open only to those who have special native equipment. "Poets are born, not made." And the same may be said of musicians, artists, and sometimes of mathematicians, chess-players, and others. Sometimes a person may be quite imbecile in most directions and yet a genius in some one direction, but these people are not sufficiently balanced to constitute a problem for educational psychology. There are other avocations which require men of average intelligence, and yet which call for no special abilities, nor yet for technical training. But the majority of the world's work is performed by people who possess different degrees of mental ability, and of educational equipment. For these intelligence is not the only criterion of success. Dominant characteristics, educational advantages, technical training, attitudinal differences, environmental conditions, economic circumstances, all enter and play significant parts in the determination of one's vocational aptitude. "It takes all kinds of people to make a world," and "the differences in kinds" means that we are able to find people who are able to do well the different things that need to be done. The vocational test attempts to measure not only intelligence, which of course is necessary, but also special aptitudes to meet specific demands called forth by the particular vocations.

One of the later uses of standardized measurements is in educational progress. I referred in the first chapter to the investigations which were carried out by Starch and Elliott showing the great disparity with which trained specialists marked so accurate a subject as geometry. Similar investigations by other workers yielded similar results. The meaning of these results is that the ordinary examination conducted along the usual lines is not a fair test of achievement, especially for comparative purposes. Here in the Madras Presidency the chairman of an examining board, besides marking his own quota of papers, has to read a certain percentage of the papers which the assistant examiners have valued, so as to guard against a multiple standard in the evaluation of a paper. This is virtually a public acknowledgment of the incompetency of the examination system as a standardized method of testing, and a genuine effort to correct the error inherent in the system. But there is no way of comparing the results of our examinations with those of Bombay, Bengal and other provinces, much less with Europe and America.

The achievement test is not intended as a measurement of intelligence, but as a measurement of the results of teaching. The intelligence test is intended to inform us about the child's capacity to learn ; the achievement test about what he has learnt. The former measures ability ; the latter measures attainment. The one tells us of possibility ; the other of actuality. The first reveals potentiality ; the second, progress. The intelligence test measures

general capacity; the achievement test measures particular attainments. The former is thus diagnostic of native skill; the latter is diagnostic of acquirements and of educational methods. On the basis of the former you may classify people for work and instruction; on the basis of the latter you can organize a school.

McCall in his recent book *How to Measure in Education* has summed up in a number of theses the true place of measurement in education. I cannot do better in closing this chapter than quote his theses—

1. "Whatever exists at all, exists in some amount"—after Thorndike.

2. Anything that exists in amount can be measured.

3. Measurement in education is in general the same as measurement in the physical sciences.

4. All measurements in the physical sciences are not perfect.

5. Measurement is indispensible to the growth of scientific education.

6. Measurement in education is broader than educational tests.

7. There are other things in education besides measurement.

8. To the extent that the pupil's initial abilities or capacities are unmeasureable, knowledge of him is impossible.

9. "To the extent that any goal of education is intangible it is worthless"—after McMurray.

10. The worth of the methods and materials of instruction is unknown until their effect is measured.

11. Measurement of achievement should precede supervision of teaching method.

12. Measurement is no recent educational fad.

13. Tests will not mechanise education or educators.

14. Tests will not produce a deadly uniformity.

CHAPTER III.

INTELLIGENCE TESTS FOR JUNIOR GRADES.

It was the desire to discover the causes of retardation which moved the Board of Education in Paris to appoint Alfred Binet to conduct his now famous research which led to his invention of a scale of intelligence tests. It soon became evident to Binet that the chief cause for retardation was defective mentality in the majority of instances. So that the original objective of the tests was to discover who among the Parisian school-children were sub-normal. The usefulness of the scale as an educational instrument was afterwards to be discovered. In the beginning it was to diagnose feeble-mindedness and other less radical cases of sub-normality that the tests were devised. But what is feeble-mindedness? Feeble-mindedness has been variously defined according to the standpoint, but one possible point of view is that of mental age. A feeble-minded person on that basis is one whose mental age is considerably below his physical age. Even though the person be an adult physically his mental age will be equivalent to that of the average child in one of the junior grades.

It is the intelligence tests for the junior grades, then, that are useful in diagnosing feeble-mindedness. That constitutes a far-reaching problem which, were we to go into it fully on its medical, economic, and social sides, would take us far afield. The Mental Deficiency Act of 1913 in England gives the following definitions:—

"The *feeble-minded* are persons in whose case there exists from birth or from an early age mental defectiveness not amounting to imbecility, yet so pronounced that they require care, supervision, and control for their own protection or for the protection of others, or, in the case of children, that they by reason of such defectiveness appear to be permanently incapable of receiving proper benefit from the instruction in ordinary schools.

"*Imbeciles* are persons in whose case there exists from birth or from an early age mental defectiveness not amounting to idiocy, yet so pronounced that they are incapable of managing themselves or their affairs, or, in the case of children, of being thought to do so.

"*Idiots* are persons so deeply defective in mind from birth, or from an early age, as to be unable to guard themselves from common physical dangers."

It will be apparent from the psychological point of view that these definitions do not define. The differentiations which are attempted are only relative, and follow no fixed standard. It would be impossible to make a classification of defectives on the basis here afforded. In all cases external control and protection are

necessary, and there is no criterion offered as to a difference in degree. It is quite conceivable that the same person might, on this basis, be classed by different examiners in all three classes. Yet these definitions represent an honest legal effort to overcome the popular vagueness in regard to these terms. I find that in one of the recent dictionaries of the English language there is no attempt to differentiate the words even to the extent that the Mental Deficiency Act does, but in the definition of one of the terms you may find the others used.

An ordinary observation of school-children will make it clear that there are degrees of mental ability. Popularly, these degrees are represented by such words and phrases as 'dull,' 'stupid,' 'bright,' 'very bright,' etc. But it is impossible for a teacher to give any accurate, standardized judgement as to the degree of brightness or dullness with which a child is characterized. Indeed it is not always that the teacher even recognizes these facts. It frequently transpires that retarded children are judged as average because they happen to do the work of the grade in which they are placed, it never occurring to the teacher that for the child to be only average in that particular grade means decided defectiveness. Terman gives several specific cases of this kind of erroneous judgement on the part of teachers. He voices the experience of all those who have had to do with the testing of children when he says that he has "often found one or more feeble-minded children in a class after the teacher has confidently asserted that there was not a single exceptionally dull child present." And he adds significantly: "In every case where there has been opportunity to follow the later school progress of such a child, the validity of the intelligence test has been fully confirmed."[1]

I have frequently had teachers say to me when the discussion of intelligence tests had begun: "I do not need any intelligence test to tell me who are the bright and the dull pupils in my class." Or sometimes: "It is a pretty poor teacher who cannot after six months with a class tell you who are the bright and who are the dull pupils." There can be no doubt of the value of the judgement of a good teacher, and the psychologist always tries to get such data, in addition to what the tests give, as the judgement of the teacher, the progress and standing of the child in school, and the environmental conditions of the child when out of school. But there are two things to remember as correctives to this popular misconception. The first is that the judgement of the teacher is formed after weeks or months of contact with the child, and observation of his work. But if the teacher had been able to administer the tests the first day that the child had entered his class, he would have been able to learn in an hour what it has required weeks or months of

[1] Op. cit., p. 24.

careful observation to teach him about the child's mentality, and probably would have much more reliable information at that. In the second place for a teacher to expect to be able to give an accurate judgement on the mentality of a child without having measured it, is analogous to a carpenter who would expect to be able to make a table of specified dimensions without a foot-rule. In each case the approximation may or may not be fairly accurate, but in either instance it is guess work.

Intelligence tests serve as correctives against all manner of vagueness and indefiniteness such as the examples given. Instead of working in the dark or with only approximations as to the meaning of words which describe mentality, we are able to give descriptions that are mathematically definite. Instead of speaking of feeble-minded children we now speak of children whose Intelligence Quotient is below 70, and these again are definitely divisible into three classes: idiots who grade roughly from 0 to 20, imbeciles who are between 20 and 40, and morons from 40 to 70. From 70 to 85 or 90 are the cases which we call slightly sub-normal or 'border-liners,' the dull, or inferiors. They cannot be classed as defectives, yet they fall distinctly below the average. There are large numbers of these children, in fact from 15 per cent to 19 per cent have an I.Q., below 90, and above 70. Moreover they have sufficient intelligence to be able to do a great number of the ordinary tasks that make up life. They need not constitute any danger to the community, especially if they are given the attention they should have in school. As we have observed, average intelligence does not mean an I.Q. of precisely 100, but it varies from 90 to 110, or some would say from 85 to 115. Again instead of using the word 'brightness' in the old vague sense in regard to a child's capacity we may say that his I.Q. is from 110 to 130. And rather than say that so-and-so is a perfect genius, we prefer to describe him as one whose I.Q. is over 130. Thus we find that the scale of measurement accompanying the tests enables us to apply to intelligence a precision comparable to the exact sciences.

It is generally conceded that the Binet scale of mental measurement has been more successful as affording a criterion for the junior than for the senior grades. There are reasons why we might expect that to be the case. In the first place the mental processes increase in complexity with chronological age, so that the earlier ages would logically be easier to test, for the simple reason that simple processes are easier to examine and measure than complex ones. The second reason is the one already mentioned, namely, that the mentally deficient whom the tests were originally intended to identify fall within the mental capacities which are judged by the junior age-grade scales. The tests were not devised with the aim of discovering children who might be super-normal, and consequently do not serve so well for that purpose. The Binet tests

or a revision of them are much the most commonly used of all the tests devised as individual tests for children of the lower grades, so we shall review them observing the phases of intelligence which they are intended to test, and the general appraisal of educationalists of their usefulness. Afterwards we may examine briefly some of the other proposed systems of measurement that have been devised.

THREE YEARS.

Binet's tests for three-year-old children were five: showing three parts of the body, repeating two digits, enumerating objects in a picture, giving the family name, and repeating a sentence of six syllables. Terman made the following changes: he increased the parts of the body from three to four, counting three out of four as correct; he added the test of naming five familiar objects demanding that three out of five be correct; he added the test of stating the child's own sex; he suggested that the sentence to be repeated be either six or seven syllables; he made the digit-repeating test an alternative, at the same time increasing it to three. Thus, Terman had six tests and one alternative as against Binet's five. Burt's revision makes the simple addition to the Binet of naming one's own sex.

The pointing to parts of the body which the examiner enumerates is a test to ascertain the child's capability of understanding simple commands. Language is psychologically an instrument for the communication of thoughts. Consequently Binet argued that the comprehension and use of language is an index to intelligibility. It assumes, to be sure, that the child and the examiner use the same language, and is of no use where there is language difficulty as in the case of a deaf child, or of a child defective in the language of the test.

The naming of familiar objects is designed to ascertain whether the child has learned to associate the names of familiar objects correctly with the objects. The association process that is here called into play is quite simple, and yet we know that it is fundamental. In adapting this test to suit Indian conditions, it is necessary to change the list of articles. The three used by Burt were a knife, a penny and a key. Terman added a watch and a lead pencil at the same time calling for only three correct responses out of five. A three-year old child may sometimes know the use of an object without knowing the name, but that does not score as correct. Hence the need of giving three chances out of five. Terman thinks to demand all correct would call for four-year ability. Miss Gordon at Saidapet substituted a quarter-anna piece for the penny but left the other articles the same. That would probably be all right for children from a community like Madras, but would include some articles unfamiliar to children in the

out lying villages. They would probably be more familiar with a slatepencil than a lead-pencil. It should not be difficult however for workers to agree upon five objects the familiarity of which would be unquestioned and to standardize the test on that basis. There are a number of workers who are now at work on the adaptation to India of the Terman revision and, they suggest the use of the following articles: key, three-pie piece, match box, glass or wax bangle, and pencil.

The enumeration of objects in pictures in the Binet scheme included three pictures, "A Dutch Home," "A River Scene," and "A Post Office."[1] The test is scored as successful if the child enumerates as many as three objects in each picture spontaneously. All that is expected of a child at three years is enumeration. If the child does more, such as a little description, that scores as correct. But description is not expected until the sixth year, whereas interpretation is not anticipated until the twelfth year. The usefulness of the test is in ascertaining the ability of the child to enumerate which involves recognition, and again implies a simple process of association.

The naming of one's sex was prescribed for the fourth year by Binet and Goddard who thought that three-year-old children could not pass it. Both Burt and Terman find it suitable for the three-year standard. The test is a simple test of discrimination.

Giving the family name is unanimously decreed to be a fair test of three-year mentality. The child will be much more familiar with his given name, but will doubtless have heard his surname quite frequently enough to know it. Of course there are some who are unable to respond to this question, but that is inevitable with any test, and constitutes the reason for having several tests for an age instead of merely one.

The repetition of a sentence involving six or seven syllables does not imply that the child should be using sentences of those dimensions in ordinary communicative processes, nor even that his power of comprehension should be so tested. He ought to be able to repeat that number of syllables, whether or not he comprehends them. A child of that age is very fond of imitating sounds and words whether it understands or not, so that it should not be difficult to secure a response. This calls for one of the simplest types of mental integration, and the fact that it is beyond the capacity of idiots and low grade imbeciles shows that it is a real test of mental ability. A very good way to conduct this test is

[1] Four pictures have been drawn which, it is believed, will be better suited to conditions in South India. They are "An Indian Home," "The Bazaar," "The Potter," and "A Street Scene." It is a pleasure to be able to produce them for the first time at the end of this volume, as figures 1, 2, 3 and 4. The Oxford University Press, Madras, is publishing them on cards for use in practical testing.

suggested by Burt. A number of words and short sentences are arranged in order of length of syllables, and the child is tested by beginning at the easier (2 syllables) and proceeding with increasing difficulty until the limit of the child's power is discovered. Burt follows the same method also in the repetition of digits.

The repetition of digits is one of the Binet tests which Terman reserves as an alternative. Here the associative process is called into function, and as Binet says: "The association of ideas triples the memory span." Binet found that three-year-old children could usually repeat two digits but few of them could repeat more. But Binet said that the digits were to be pronounced at the rate of two per second. Terman finds that the three-year-old child can repeat three digits, but says that two per second is too fast. Just a little faster than one per second is the proper rate. The plan is, as with the syllables, to begin with the pronunciation of two syllables and increase the number until one has ascertained the limit of the child's capacity.

These are the tests for three-year-old intelligence. It will be necessary to experiment with Indian children to see whether the same standard will suffice. The Saidapet experiments tend to show that this group of tests measures four or five-year-old intelligence for children here. Of course the lower class in the Model School was composed of children of that age, and the lowest test available was that for the third year, so it was natural to use it for these children.

FOUR YEARS.

Binet has but four tests for four-year mentality. These are the giving of one's sex, the naming of three familiar objects, the repetition of three digits and the comparison of two lines. Burt's tests for this age are the repetition of six syllables, the repetition of digits, the counting of four coins, the comparison of faces, and the comparison of lines. Terman's revision includes the comparison of lines, the discrimination of forms, the counting of four coins, the copying of a square, a simple test in comprehension, the repetition of four digits, and an alternative test of the repetition of twelve to thirteen syllables.

It will be seen that three out of four of Binet's tests have been moved up to the third year by Terman. Burt agrees with Binet in keeping the digit repeating test for three digits as a four-year test.

The comparison of lines is a test used by all three men. It is the simple test of telling which of two lines that have been shown is the longer. In the Terman revision there are three pairs of unequal lines shown and the child is expected to make three correct responses. No hesitation is permitted. Binet found this a good test for eliminating the feeble-minded, because an imbecile

who would shut the door when the command was accompanied with a gesture, but could not do so without the gesture, always failed on this test. It is a test of comprehension, discrimination and comparison, all fundamental, yet here presented in an elementary form. It is however more often a test of language comprehension than of actual discrimination, for a child who would unerringly choose the larger of two pieces of biscuit or sweetmeat sometimes fails in the test.

Terman introduces a test for this year in the discrimination of forms.[1] Two sheets of paper each contain ten forms, exactly alike. These are : elipse, square, triangle, circle, rhombus, rectangle, octagon, cross, and three irregularly formed figures. The examiner places his finger on a figure on one card and asks the child to point to the corresponding figure on the other card. The test was devised by Kuhlmann who standardized it at seven correct responses out of ten. The test is not unlike the form-board test, and tries the child's power of discrimination a little more than the comparison of lines. It also tests the attentiveness of the child, as well as his visual perception of form. It is a question for investigation how well this would test the intelligence of children from the backward classes and the outlying villages of India. The training of the sensory mechanisms as in observation is undeniably a great help in making responses of this type, and the social environment from which the child comes is largely determinant of the amount of such training that he may have had.

The counting of four coins is used as a four-year-test by Burt and Terman, though Binet, Kuhlmann and Goddard used it for a five-year test. It has been objected that this test implies a certain amount of instruction rather than intelligence against which objection Binet urges, " Where is the being so deprived of tutelage that no one has ever taught him to count ? " He even found that all imbeciles with sufficient intelligence had learned to count. The test does not demand a mastery of numbers or an analysis of calculation, yet experience with it shows that success does not depend on schooling, for most children succeed before they have had any such opportunity. The quarter-anna coin is the one being used in India.

The copying of a square is allotted by Binet, Burt, Goddard and Kuhlmann to the fifth year, though Terman places it in the earlier year, and other workers have found that it correlates well with the other tests with which Terman has grouped it. The child is simply asked to copy the square on a paper. Binet had it done with pen and ink, but the American revisers prefer the pencil. The test is passed if, in one out of three attempts, the child

[1] See Fig. 5 at the end of this volume.

produces a drawing that is recognisable as an honest attempt to reproduce the square. This is a good test to illustrate the three points which Binet made about the psychological value of the tests. The printed square serves as the suggestion of the end to be achieved, and after the child has drawn the three copies his auto-criticism is called into play by asking him to tell which of the three he considers the best. Sub-normals invariably lack in this ability, and of course very young children show the same deficiency. Probably the reason that Binet places the test a year in advance of Terman is because he demands the use of a pen which is obviously more difficult. But as a test of intelligence it is a questionable procedure to introduce that element, facility in which demands practise rather than intelligence.

The comprehension test consists of asking the child such simple questions as: "What must you do when you are hungry?" "What ought you to do when you are cold?" "What should you do when you are sleepy?" Twenty seconds may be allowed in which to answer each question. The questions are intended to elicit responses of a sufficient degree of pertinency to show that the child comprehends the meaning not only of the words but of the situations. Terman rightly remarks that "it probably requires more intelligence to tell what ought to be done in a situation which has to be imagined than to do the right thing when the real situation is encountered." With this test two correct responses are demanded out of three.

The digit-repeating test is used again. Binet considered that a four-year-old child should be able to repeat three digits. Burt agreed with him. Terman found that 75 per cent of four-year-old children could repeat four digits, if they were pronounced slowly so that nearly four seconds were consumed in the pronunciation. Out of three series the child is expected to pass one correctly.

The syllable-repeating test comes in again. It is rather surprising to find such a wide divergence between the success demanded by Burt and Terman for this age, Burt placing the number at six and Terman at twelve to thirteen. Three sentences of that length are given, and the repetition of one of them correctly is scored as a success. In the syllable-repeating tests for the younger children no examiner pays any attention to defects of pronunciation due to imperfect development in the use of language.

Burt includes in this year a test in the comparison of faces. All the investigators use the six faces[1] which Binet first used, showing them to the child in pairs, and asking the child in each case which is the prettier of the two. Terman placed this test in the five-year series, and Binet in the six-year-old. It is better to use the same faces as Binet since the comparisons have been so well standardized.

[1] See Fig. 6 at the end of this volume.

The aesthetic attitude is one that appears very early in life and depends upon natural tendencies. This test is interesting as a criterion of the age at which the ability to make aesthetic comparison develops. All of the workers agree that the development, if it is not a phase of intelligence itself, at least develops parallel to intelligence. Moreover tests of the feeble-minded lead to a substantiation of the parallel development of aesthetic judgement and intelligence. Imbeciles of four-year age mentality, though their chronological age be forty, have no chance of passing the test, according to Terman. The children tested at Saidapet led to the conclusion that the test is rather difficult even for six and seven-year-olds. Undoubtedly environmental conditions would alter the situation here. Children from superior surroundings who have frequently heard adults admire the beautiful and decry the ugly must develop earlier the aesthetic judgement than children who come from environments where little attention is paid to these distinctions.

FIVE YEARS.

Binet's list for five-year-old tests includes the comparison of two weights, the copying of a square, the repetition of a sentence of ten syllables, the counting of four coins, and the game of patience with two pieces. Burt's adaptation includes the performance of three commissions, the copying of a square, the repetition of ten syllables, the giving of one's age, distinguishing morning from afternoon, naming the four primary colours, the comparison of two weights, and giving the number of one's fingers. Terman approximates more to Burt than Binet. He includes the comparison of weights, naming the colours, the execution of three simple commissions, giving one's age (alterative test), the game of patience, and the aesthetic comparison.

The comparison of weights, it is agreed by all three, is a test suitable for this age. The two weights should be identical in external appearance, size and shape, but must differ radically in weight. Three and fifteen-gram weights are frequently employed. The child is asked to try them and tell the instructor which is the heavier. The relative positions are changed and the child is asked three times to respond. Two successes in three trials score as correct. This test is decidedly more difficult than the comparison of lines. The visual perception which the former calls for comes into operation earlier in experience than muscular discrimination for which this test calls. The test has marked psychological value. It involves first, comprehending the fact that the weights of the two boxes are to be compared; second, the ability to hold instructions before consciousness long enough to make the comparison; third, the conative ability to concentrate attention and overcome abstractions; and fourth, the appreciation

of difference in weights. The imbecile often starts off as though he were going to perform this test according to instructions, but ends by playing with the two weights instead of trying to compare them.

The naming of the four primary colours from four well saturated colour cards is a test which Terman and Burt use for the fifth year. Goddard placed it in the seventh year in agreement with Binet's 1911 revision of his own original in which he had it in the eighth year. Several other investigators place it in year five. It is as Binet said a test of "the verbalization of colour perception." It indicates whether or not the child can associate the names of these colours correctly in perceptual processes. To be sure it would not succeed in a case of colour-blindness, but colour-blindness is not an indication of defective mentality. But in case of children with normal visual power it is a good test of the visual discrimination of colours. Like the aesthetic comparison test, it is somewhat more largely influenced by environmental conditions than many other tests. Girls are found to do better than boys, on the average.

The execution of three simple commissions is placed by both Burt and Terman in the five-year scale. The three commissions are named together: "Do you see this key? Go and put it on the table there. Then shut the door. And after that bring me the book on the chair near the door. Do you understand? First put the key on the table, then shut the door, then bring me the book." All three commissions must be executed without prompting and in the given order to score success. Success depends on the ability to comprehend and then to carry out the instructions. It is the test of a type of response which in actual life we are constantly called upon to make, a response that depends on intelligent comprehension and memory. There are many people of defective mentality who can be entrusted with one commission, but who are quite at sea when given more. Environmental conditions where there is a fair degree of co-operation and discipline would no doubt minister to the success of this test.

Giving one's own age is adopted by both Burt and Terman as a test suited to the five-year-old level of intelligence, though the latter does not value it very highly and uses it only as an alternative. He says however: "If the child has arrived at the age of 7 or 8 years and has had anything like a normal social environment, failure in the test is an extremely unfavourable sign." As for its psychological importance it gives evidence of little more than a normal interest in life and a memory process. Most normal children do remember their age, whereas middle-grade imbeciles even in advanced years do not. In India there is not the same custom of celebrating birthday anniversaries which prevails in the West, and hence investigators here find the test rather unsatisfactory. Miss Gordon had one girl tell her with perfect assurance

that she was 15 years, and a few minutes later with just as much assurance that she was two months old. There was a boy of about eleven years in the Kurnool school when I was there who used to give his age as 35 with no apparent appreciation of the absurdity. Binet first made use of the test for six years and afterwards dropped it entirely.

Terman introduced the test of giving definitions in terms of use. The words suggested were: *chair, horse, fork, doll, pencil,* and *table.*[1] The procedure was as follows: "You have seen a chair. You know what a chair is. Tell me, what is a chair?" Binet, followed by Burt, placed this test in the sixth year, but most investigators agree with Terman. The words selected must, of course, in the case of such young children be concrete, so that a functional definition is possible. The defining process demands a higher process than simply knowing a thing, and this test is intended to test that knowledge—a part of the apperceptive process. It is possible to classify the degrees of precision in definition quite minutely. But the concern here is to secure the simplest kind in terms of use.

Binet designated as the game of patience the test which Terman also adopted as a test for five-year mentality. Two rectangular cards are taken, each 2 × 3 inches, one of which is divided into two triangular pieces by cutting along one of the diagonals. The child is invited to take the two triangular pieces and so put them together that they will exactly resemble the rectangular piece. Binet believed that this test affords an excellent illustration of the psychological processes involved in intelligence, namely first, keeping in mind the end to be attained; second, trying various combinations with the end in mind; and third, auto-criticism of the attempts made. He called the test a "test of patience" because it requires a certain degree of persistence for successful solution. He also pointed out that various complications of the game can be worked out, so that the more complex would try the skill even of adults.

SIX YEARS.

The Binet tests for six years include distinguishing morning from afternoon, definition in terms of use, the copying of a lozenge, the counting of thirteen coins, and the aesthetic comparison test. Terman agrees as to the distinguishing of morning from afternoon, while Burt would place it a year earlier. All three agree in including the counting of thirteen coins. Burt and Terman include distinguishing between right and left. Both of them have the

[1] The following words are suggested for definition by Indian children: chair or stool, baby, ball, horse, water-pot, hoe, and table.

naming of four coins. Terman has also these tests:—finding omissions in pictures, second degree of comprehension, and the repetition of sixteen to eighteen syllables. Burt has these:—drawing a diamond or rhombus from copy, transcription of three words, naming the days of the week, the patience game, the definitions in terms of use, the repetition of five digits, and the simple description of pictures.

The discrimination between morning and afternoon is a simple test in the perception of temporal relations. Terman thinks that certain perceptions of spatial relations come earlier. It is of interest to observe the development of the child in ability to make such distinctions. Binet remarked on the ridiculousness of a programme which he had found operative in some schools, where they were actually trying to teach the rudiments of national history to children who had not learnt to distinguish between forenoon and afternoon. Terman rightly points out two weaknesses in the test—(i) the language difficulty—some children may be able to appreciate the distinction before they can do so verbally; (ii) the play of chance—at least fifty per cent would be right by guess work.

The copying of a diamond was introduced by Binet for the sixth year. It was his experience with imbeciles that those who were able to copy a square failed in the attempt with a diamond. And children at five who could copy a square failed in their attempts with a diamond. It demands a little more advanced piece of perception, and the diamond is a bit more difficult of reproduction. Binet placed it in the sixth year, at the same time acknowledging that only half of six-year-olds could do it. Terman puts it in year seven.

Counting thirteen coins is a test of six-year intelligence. It has been suggested that it tests instruction rather than intelligence, but the general opinion of investigators is to the contrary. By the age of six a normal child should evince enough interest in affairs to have learned spontaneously to count up to thirteen numbers. Only an exceedingly unpropitious social environment would fail to inspire that amount of native interest. Binet cites three conditions requisite for a successful test: (i) the child must be able to count to thirteen; (ii) the child must touch each coin separately and name the corresponding number which demands intelligent guidance since the tendency is for the hand to run in advance of the tongue; (iii) the child must neither forget any coin nor count any the second time, which involves the use of a discriminating method. Feeble-minded adults of the five-year level of intelligence cannot be taught to count to thirteen without much laborious instruction.

Distinguishing right from left is placed by Binet at seven years, but Burt and Terman put it in the six-year group. The test

is administered thus: "Show me your right hand." "Show me your left ear." "Show me your right eye." The test may be once repeated. Five out of six responses must be correct to score success. This is a test of spatial orientation, of which other tests might be given, such as up and down, far and near, before and behind, etc. But the test suggested has been standardized, so that results can be compared better than in other cases. Bobertag found that these other distinctions were mastered earlier than the right and left distinction, a matter for which there are several possible explanations: frequency with which the words are heard, frequency with which the distinctions are called for, differences of the orientation demanded, variations in the kinaesthetic sensibility called into play, associative connections, etc. Many people learn to make the distinction between right and left by means of an association, so that with such people the test becomes a test of association as well as of discrimination. One little girl according to Terman responded by trying to wink first one eye and then the other, explaining herself by saying that she knew that she could wink her left eye but not her right.

Terman and Burt both include the test of naming four coins[1] for this age, the test being passed if three out of four are correctly named. Binet gave the test a place in his 1908 scale for the year seven, but omitted it from the 1911 scale. Goddard also omitted it from his adaptation. Some have criticized the test as depending on instruction rather than intelligence, but its defenders claim that failure to learn the names of the common coins by six years betrays a lack of spontaneity of interest which does not depend on schooling. Statistics show that American children from poorer homes do slightly better than those from homes of wealth, while the tendency among Indian children seems to be without regard to such distinction of environment, for all to be able to respond correctly.

Finding omissions in pictures is made a test of seven-year-old mentality by Binet and Burt, but Terman and others put it in the sixth year. Four pictures[2] are shown to the child, in one case the eye is missing, in another the nose, in another the mouth, and in the fourth the arms. The child is asked to indicate which features are missing from each picture. It is one of the so called "completion tests" that from the given parts of a whole call for the recognition of what is missing. The "whole" may be a picture, as in this case, or a story, or a sentence. Whipple in his Manual has a good discussion of the completion method.[3] Ebbinghaus investigated the method very carefully and the result

[1] The coins used in India are: anna, quarter-anna, rupee, and two-anna (nickel).

[2] See Fig. 7 at the end of the book.

[3] Vol. II, pp. 649-666.

of his investigation showed a very marked positive correlation between success in this test and general ability. This particular form of the completion test calls for the most elementary type of ability in recognition of omissions. It requires a visual perception of form sufficient to attain a coherent idea. Many feeble-minded individuals have great difficulty with tests of this type.

Comprehension in the second degree is tested by these three questions: (*a*) "What is the thing to do if it is raining when you start to school?" (*b*) "What is the thing to do, if you find that your house is on fire?" (*c*) "What is the thing to do if you are going some place and miss your train?"[1] These questions demand a more developed type of comprehension than those which were used in the four-year tests, and consequently a greater variety of correct answers is possible. Binet's experience with French children was that very few children could answer them at six years, at seven and eight years half could answer, at nine three-quarters, and at ten all.

SEVEN YEARS.

Binet's tests for seven years are distinguishing between right and left, description of a picture, the execution of three commissions, counting nine sous (three single and three double), and naming the four primary colours. The Terman revision includes giving the number of fingers, the picture-description test, the repetition of five digits, tying a bow-knot, giving differences from memory, copying a diamond, and naming the days of the week. Burt's revision includes the recognition of missing features in pictures, counting three pennies and three half-pennies, stating differences between concrete objects, the repetition of sixteen syllables, and writing from dictation.

The picture-description test demands a little greater ability than the mere enumeration called for when the same pictures are shown to three-year-olds. The correct response depends somewhat on the way in which the question is asked. It must not be so put as to call for mere enumeration. Here again, owing to the increase in complexity of the mental processes with advancing age, the variety of possible correct answers increases.

The sous-counting test was used by Binet, and Burt substituted pence and half-pence for sous and two-sous pieces. In America there is no two-cent coin, so Goddard substituted one and two cent postage stamps. Terman omits the test, perhaps because stamps are less familiar than coins which militates against the usefulness of the test. The test calls for discrimination between

[1] The following questions have been substituted for the second and third questions of Binet: (*b*) What is the thing to do, if your brother falls into a well? (*c*) What is the thing to do if you are sent to buy a cocoanut and lose your money?

the two values, as well as the ability to add correctly, whether the addition is done by ones or the double coin is counted as two. Terman has substituted the test of counting fingers which calls for the same spontaneous interest in numbers. Not many children seem able to remember the number of fingers which they have unless they count them, and the same is true of the feeble-minded.

Tying a bow-knot is a new type of test, more of the performance type. The child is shown a bow-knot made by tying a shoe-string around a stick and is given a minute in which to tie another shoe-string into a bow. Terman says that the fact that children of more advanced chronological age but of seven-year mental age do not succeed any better than children who are young physically, indicates that it is a good mental test. Environment and instruction may tell against the test, and girls succeed somewhat better than boys. But these factors are not as prominent as might be anticipated. The test calls for skill in the direction of play impulses and in ordinary motor control, interest in common objects, and the ability to form correct associations with their accompanying motor reactions. The bow-knot is not used as frequently in India as in the West, and consequently the Indian children do not do well in this test. Miss Gordon suggests the substitution of the bow line which is more commonly used in the Indian house-hold.

Stating differences between concrete objects from memory is placed by Terman and Burt in the seven-year group. Binet places it in the eight-year group although he acknowledges that most children at seven pass it. Goddard found 97 per cent pass it at eight years, and Dougherty 90 per cent at six years. Three comparisons are called for: a *fly* and a *butterfly*, a *stone* and an *egg*, and *wood* and *glass*. In each case the child must discover and state the difference without hint or suggestion. The investigators are agreed in approval of the test because schooling plays such an insignificant part in determining the child's response. It tests a higher type of mental process perhaps than any of the tests discussed thus far, the process of contrasting differences which involves associative processes more complex than simple similarities. Association by contrast depends on there being a fundamental likeness to begin with, and the meaning of the difference depends upon the primary likeness. In the test, the difficulty is increased by the fact that the objects to be compared are not present to the senses, so that the comparison depends upon memory images. There are, of course, a considerable number of possible correct responses, and the manuals give many suggestions for scoring on the basis of satisfactory and unsatisfactory. But one thing must be guarded against, namely stereotyped answers to all three which would indicate an absence of intelligent thinking, even though they might happen to be right in a specific feature.

Naming the days of the week is defended by Terman as another kind of time orientation which an intelligent child readily learns to make. In some cases the correct response may be due to rote memory, but 'checking-up' questions will make that matter clear. Miss Gordon reports an interesting type of association obviously due to wrong instruction. Some of her subjects named the days of the week correctly, but without stopping to take breath continued to enumerate the names of the months and concluded with saying that there are 7 days in a week, and 12 months in a year.

The repetition of digits in reverse order was first suggested by Bobertag in 1911. Subjects cannot repeat as many digits in the reverse order as in direct order. Children at seven can repeat five in direct order, but only three in reverse order. As a test of intelligence, repetition in reverse order calls into play more conscious attention and depends less upon mechanical associations or pure memory. Feeble-minded children find it a most difficult test on that account. More intelligent subjects usually adopt a method of grouping, more frequently into twos, and are thus able to repeat a larger number. The test is fundamental because its success depends on ability in manipulating images, and the manipulation of images in consciousness is the mechanism of the thought processes.

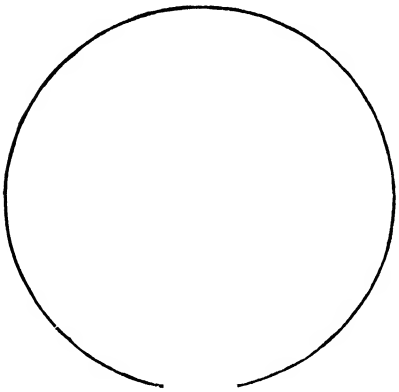
EIGHT YEARS.

The Binet tests for eight-year mentality are comparison of pairs of remembered objects, counting from 20 to 1, indicating omissions in pictures, giving the day of the week and the date, and the repetition of five digits. The Burt revision has a reading and reproduction test, answering easy questions, i.e., comprehension tests, counting from 20 to 0, giving the full date, and making change so as to show knowledge of the coinage of one's own country. Terman's revision includes the inferior plan of the balland-field test, the counting backwards (20 to 1) test, the comprehension test (third degree), giving definitions superior to use, the vocabulary test (20 definitions), and two alternative tests, viz., naming six coins and writing from dictation.

Counting from 20 to 1 involves certain processes of which we have already taken note in the repetition of digits reverse order with the addition of a memory process. One must be able to count first from 1 to 20 before he can reverse the process. In addition to memory there is required a comprehension of the relative numerical values, sustained attention until finished, an association which is recalled in reverse order to the order in which it was formed, and a conscious end towards which the child persists. Yerkes suggests that the experimenter count from 25 to 21 to give the child the idea, Binet and Terman suggests counting 20–19–18, and asking the

child to continue. One error is permitted. The investigators differ as to the time allowed from 20 to 40 seconds.

Terman introduces the ball-and-field test here. The procedure is to draw a circle, about two and one-half inches in diameter leaving a small gap: Then say to the child: "Let us suppose that your ball has been lost in this round field. You have no idea what part of the field it is in. You do not know from what direction it came, how it got there, or with what force it came. All that you know is that the ball is lost somewhere in the field. Now, take this pencil and mark a path to show how you would hunt for the ball so as to be sure not to miss it. Begin at the gate, and show me what path you would take." The responses to this test have been classified by Terman into four groups: (i) failures to comprehend what is wanted; (ii) the search carried out with no definite plan; (iii) the inferior plan which is declared satisfactory at age eight, a common characteristic of which is the tendency to make lines more or less parallel; (iv) the superior plan which is satisfactory for a twelve-year test, which may be concentric circles, a spiral or parallel lines joined at the ends. The test, being of the performance type, calls for practical judgement and adjustment, overcoming to some extent the excessive language stress of the Binet scale.



The comprehension test, third degree, calls for the same type of response as the previous comprehension tests only that it is slightly more advanced. The questions suggested are three—

(*a*) "What is the thing to do when you have broken something which belongs to some one else?"

(*b*) "What is the thing for you to do when you notice on your way to school that you are in danger of being late?"[1]

(*c*) "What is the thing for you to do if your playmate hits you without meaning to?"

Binet used this test for the tenth year and in this he was followed by Goddard, but the Stanford data and Burt's data indicate that it belongs rather to the eight-year level. Binet thought that the comprehension called forth in such questions was in some respects

---

1 For the second question the following has been substituted by workers in South India: "What is the thing for you to do, if you see a buffalo in some one else's paddy-field?"

a better test of intelligence than any of the previously mentioned ones.

The test of giving similarities calls for an expression of one of the elementary forms of association. The objects to be compared are: an *apple* and a *peach*, *iron* and *silver*, a *ship* and an *automobile*, and *wood* and *coal*.[1] The child often tends to err in stating differences rather than likenesses which seems to be an easier type of mental process. That point comes out especially with the sub-normals who persist in giving differences even after reproved for so doing. "The more essential the resemblance," says Terman, "the better indication it is of intelligence."[2] Of course the test involves things that have fundamental similarities, so that a correct answer does not call for any conundrum-solving ingenuity but for a normal mental process. Two out of four correct responses are scored as successful.

Giving definitions superior to use calls for a response a little more advanced than the fifth year test. It may be descriptive, may define in terms of component parts, or may classify the object and give its relationship. The shades of differentiation which are evoked are good indications of the development which the child's intelligence has attained. We observed in the second chapter that what marks the intelligence of the human from that of the lower animal is the ability to abstract and form concepts, and this test often gives an insight into the rudimentary forms of this process in the child's consciousness. Terman's words are *balloon*, *tiger*, *football* and *soldier*. The substitution of *ship* for *balloon*, and of *kite* for *football* is suggested for India.

The vocabulary test introduces us to something new, and its standardization has meant a great deal of arduous labour on the part of the psychologists. A list of one hundred words is given in the record booklet of the Stanford revision. The object is to ascertain how many of the words the child is able to define, the words being arranged in their order of approximate difficulty. A scale has been arranged on the results of testing many hundreds of children, which is as follows:—

Children of	8	years	...	...	...	20	words.
„	10	„	...	...	...	30	„
„	12	„	...	...	...	40	„
„	14	„	...	...	...	50	„
Average adult			...	...	...	65	„
Superior adult			...	...	...	75	„

The list of 100 words was made by a selection according to careful planning from a dictionary of 18,000 words. On that

[1] The following objects are suggested as suited to Indian conditions: *wood* and *bròttìes*, *mango* and *orange*, *iron* and *silver*, *train* and *jutka*.

[2] Op. cit., p. 219.

reckoning it is calculated that multiplying the number of correct definitions which the child is able to give by 180 will give the approximate size of his vocabulary. Thus a child who correctly defines 20 words has a vocabulary of 20 × 180 = 3,600 words, one who defines 30 words will have a vocabulary of 5,400 words, 50 definitions for 9,000 words, 75 definitions for 13,500 words, etc. The test is designed to discover the range of ideas which the person possesses rather than to measure his ability in exact definitions. If a child can give one of the meanings of a word with fair correctness it is scored as a success. The vocabulary test was arranged and standardized by Terman and Childs in 1911, and has proven to be of higher value as a test of intelligence, according to the former, than any other test in the Stanford revision. The feeble-minded find it an exceedingly difficult examination, very frequently offering definitions with no sense or significance for words the meaning of which they do not know. It will be a task here in India to arrange lists of words for the various vernaculars that will be standardized and afford some criterion paralleling that of the Stanford list. Some work is being done by workers in the Tamil, Telugu, and Hindi language areas, but much more needs to be done in these and other areas.

NINE YEARS.

Nine-year intelligence was tested by Binet with the following tests: giving change, definitions superior to use, recognition of coins, enumeration of months, and comprehending simple questions. Burt's revision includes the repetition of six numbers, enumeration of the months, recognizing coins, reading and reproduction, and definitions superior to use. The Stanford tests for the age are giving the date, arranging five weights, making change, repeating four digits reversed, using three words in a sentence, finding rhymes, and two alternative tests of enumerating the months, and counting the value of stamps.

Giving the date is an indication of time orientation a little more difficult than what we have had because it involves the divisions of the year, the month and the week. Binet and Bobertag found that children experienced more difficulty in naming the year than any of the parts of it, but Terman found that in his experience the children realized the parts of the tests as of equal difficulty.

Discrimination in weights where there are five weights to be considered demands quite a good deal finer type of discrimination than where there are only two to be compared as in the fifth year test. The weights suggested are 3, 6, 9, 12 and 15 grams, though Kuhlmann used 3, 9, 18, 27, 36 and 45 grams. The greater the difference in the weights the easier the discrimination. The psychological elements that are involved are realisation of the end, comprehension of the task, an appropriate choice of means to

the end, and persistence of effort. These are all elementary mental processes which are being constantly demanded in actual experiences: so that success in the test is a good indication of the functioning of normal processes of intelligence. The possibility of failure is more varied than in some of the earlier tests, and it is wise to record the cause for failure, as that too is significant. It may be due to lack of comprehension, or to inadequate methods, or to lack of perseverance. One advantage which the test has is that it is a manipulation test, depending less upon the use of language for success than many of the other tests. It gives us information not only about mental processes, but also about their motor concomitants, and tests which call for motor as well as mental elements are invariably of more interest to the child.

Making change was placed by Burt in the eighth year, but Binet and Terman place it in the ninth. The problem is solved theoretically rather than practically, because coins are not used, neither is the child allowed the use of pencil and paper. It will be better to state the three problems as they were adapted to the Saidapet experiments, since the difference in the coinage must be observed. Naturally Binet used French, Burt English and Terman American coins. These are the problems:

(*a*) "If I were to buy four rupees worth of mangoes and give the bazaar-man a ten-rupee note, how much would he give me back?"

(*b*) "If I bought As. 10 worth of sweetmeats and gave the bazaar-man a one-rupee note, how much would I get back?"

(*c*) "If I bought eight annas worth of rice, and gave the bazaar-man a five-rupee note, how much would I get back?" There is some difference of experience among the investigators as to the correct age in which to place these tests. In Saidapet it was found that the tests were too easy for the age, and could be done by all children of seven and eight years. The test involves comprehension of the nature of the problem, and a choice of the correct mode for its solution. Many defectives are unable to handle this type of problem, because it calls for something more than routine which seems to be all that they can master.

The use of three words in a sentence is a type of test which now appears for the first time. Three problems of this type are given. The words[1] used by different investigators vary, as

(*a*) boy, ball, river ... ... Terman.
(*b*) work, money, men ... do.
(*c*) desert, rivers, lakes ... do.
(*d*) London, river, money ... Burt.
(*e*) Paris, river, fortune ... Binet.

[1] The words suggested for India are those marked (*a*) and (*b*) to which are added: jungle, rivers, tanks.

The student is then asked to compose a sentence in which all three words are used. The European investigators conduct the test with pen and paper, but the American orally. It is known as the "Masselon experiment" after the man who devised it. Success is attained if the pupil composes a sentence that makes sense in either simple or compound form with not more than two distinct ideas. The experiment tests the child's ability to form logical associations on the basis of which he can make definite assertions. A dull child may sometimes succeed in expressing a sentence devoid of logical absurdity, and yet containing two rather disjointed remarks. One of the marks of mental sub-normality is poverty of associations, and this test is well adapted to bring out any such defect. Brighter intelligence is characterized by richness of associations, and such tests give a criterion to that in the speed and logical correctness with which the child responds.

Finding rhymes is another test that draws on the associative tendencies of the child. A sample is given to the child such as cat, hat, rat, mat, sat, etc. Then the child is given one minute each for three words in which to name as many rhyming words as possible. These words[1] are *day, mill, spring.* The type of association here called for is auditory similarities. To find rhymes for a given word demands a process of exploration among the verbal associations, always remembering the dominant interest in sound likenesses. Many associations may come to the child, but he must inhibit those that are irrelevant and select the relevant for success. It is more than a pure vocabulary test, for many sub-normals may have quite sufficient associations, and yet fail for lack both of inhibitory and selective abilities. There are certain data which prove beyond a doubt the efficacy of the test as one of intelligence. Fatigue decreases adeptness in the rhyme-finding process. A person of 30 years chronological and 12 years mental age does not do as well as one of 12 years chronological and the same mental age. A nine-year-old child with ten-year mentality is invariably adept, and a nine-year-old child with eight-year mentality is invariably sluggish in the performance. The placing of the test varies with the difficulty of the words employed, Binet, using much harder words, having placed it in the fifteen-year series.

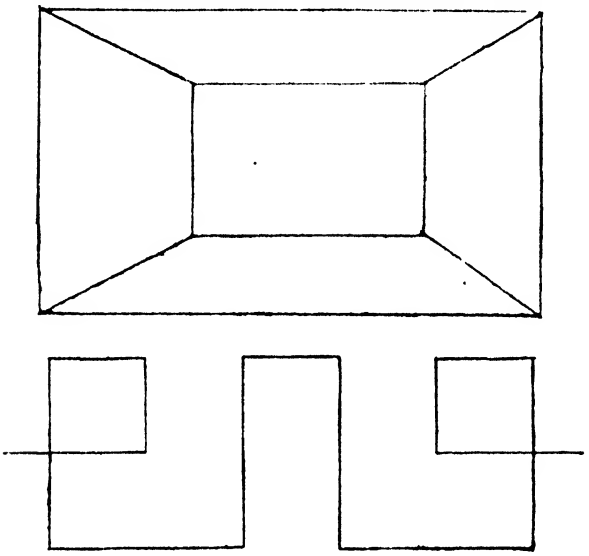
TEN YEARS.

Binet employed the following tests for ten-year mentality: discrimination of five weights, copying drawings from memory, criticism of absurd statements, comprehension of difficult questions, and using three words in a sentence. Burt uses the discrimination of weights, sentence building including three given words, and

[1] It will be necessary to adapt this test in the various vernaculars.

drawing designs from memory. Terman has the vocabulary test, thirty definitions with an equivalent vocabulary of 5,400 words being called for, the detection of absurdities, drawing designs from memory, reading and reproduction for eight memories, comprehension of the fourth degree, naming sixty words within the space of three minutes, and as alternatives the enumeration of six digits, and a form-board construction puzzle.

Binet suggested the two designs which all the investigators have adopted, and which the child is shown for ten seconds, and then asked to reproduce from memory. The test is passed if one is reproduced correctly and the other one half correctly. Neatness of execution does not count. A fair degree of exactness is all that is demanded. Binet's estimate of the value of the test is that it involves "attention, visual memory, and a little analysis." Certainly all of these elements are demanded. Without close attention the task could never be performed, and the child usually attends to the figure to the left (they are shown side by side) more closely than to the one to the right. Perhaps a child whose language is Urdu which runs from right to left would attend to the one to the right more closely and so reproduce it more faithfully. Visual memory is obviously demanded, and without it the test would be an utter failure. Analytical ability is important, for the figures are sufficiently complex for the child to be unable to reproduce them correctly unless he has grasped the various lines in some synthetic relationship. Terman's remarks are worthy of attention: "Ability to pass the test indicates the presence, in a definite amount, of the tendency for the contents of consciousness to fuse into a meaningful whole. Failure indicates that the elements have maintained their unitary character or have fused inadequately."[1] Previous training in drawing, especially from memory would undoubtedly facilitate success in performance, as some investigators report.



[1] The Measurement of Intelligence, p. 261.

The detection of absurdities was originally designed by Binet as a test of judgement, but his conclusion was that it was not a success for that purpose but tested rather timidity, deference, confidence and automatism. At first he did not announce that the statement was absurd, and was greeted by ironical laughter; later he announced that it contained an absurdity, and asked the child to point it out. With this change of method in procedure the feelings of deference, timidity, or reserve which hitherto paralyzed the judgement were removed. The original Binet absurd statements were as follows:

(i) "An unfortunate bicycle rider fell on his head and was killed instantly; he was taken to the hospital and they fear he will not recover."

(ii) "I have three brothers—Paul, Ernest and myself."

(iii) "The body of an unfortunate girl, cut into eighteen pieces, was found yesterday on the fortifications. It is believed that she killed herself."

(iv) "There was a railroad accident yesterday, but it was not a bad one; the number of dead is only forty-eight."

(v) Someone said: "If I should ever grow desperate and kill myself, I will not choose Friday, because Friday is an unlucky day and always brings unhappiness."

Terman substituted for the second and fifth above the following:

(vi) A man said: "I know a road from my house to the city which is downhill all the way to the city, and downhill all the way back home."

(vii) "An engineer said that the more cars he had on his train the faster he could go."

One unacquainted with psychological tests is likely to think the test to be as absurd as the statement it contains, on first hearing one of these absurd statements. But as a matter of fact it has proven to be one of the most reliable tests devised.[1] The detection of the absurdity calls for a type of comprehension and criticism which the backward person lacks. Without the ability to criticize the person will fail to find anything absurd in the statements, and listen to them without a protest. Binet found one difficulty with the test, namely, that many children were unable to give a clear verbal expression of the absurdity, sometimes contenting themselves with a mere repetition of that phrase in the statement which contains the absurdity. A further question is then required to encourage the child's critical ability.

---

[1] Dr. P. B. Ballard, in Chapter VII of *Group Tests of Intelligence*, has an excellent discussion of the absurdities test. It has been suggested that for India absurdities (vii), (iii), and (i) above be used, and the following two added:

(viii) I am now older than my mother

(ix) A sign says: Eleven miles to the village; if you cannot read, ask the bazaar-man.

Binet originated the test of reading and reproduction for memories, but afterwards omitted it from his revised scale, as did Goddard and Kuhlmann also. Terman introduced it in the Stanford revision for ten-year mentality. When Binet rejected it he did so on the ground that it was too difficult, but he was trying it as an eight-year test, whereas Terman uses it for ten years, quite a different matter. The child is handed the following selection and asked to read it as well as possible:

"New York, September 5th. A fire last night burned three houses near the centre of the city. It took some time to put it out. The loss was fifty thousand dollars, and seventeen families lost their homes. In saving a girl, who was asleep in bed, a fireman was burned on the hands."

After the child has read the selection and attention has been given to the reading, he is asked to report that which he has read, and each phrase which he is able to reproduce correctly is scored as a memory. Obviously this test depends a great deal on schooling, and for that reason it has been rejected by some. But there are few children at ten years who have not had sufficient opportunity to be able to make such a response as this calls for. The validity of the test depends however on the child having had normal educational opportunities so that in case of failure it is necessary to inquire into that matter. The development of mastery in language is a concomitant of the development of conceptual processes, and on that ground the test is defended by Terman as a legitimate test of intelligence. Success in performance of this test means the functioning of associative tendencies which are fundamental to recognition and reproduction.

The next test is one of naming as many words as one can in three minutes. The words must be separate, and must not be connected as in sentences or in counting. At the same time the richness of one's associations will be reflected in the test, the child of high mentality tending to make his response in the form of groups of words representing associations which readily reinstate themselves. Advancing mentality is indicated by a larger number of abstractions. Terman employs a useful analogy to describe the distinction which the test discloses. He says: "The young or retarded subject fishes in the ocean of his vocabulary with a single hook, so to speak. He brings up each time only one word. The subject endowed with superior intelligence employs a net (the idea of a class, for example) and brings up a half-dozen words or more. The latter accomplishes a greater amount with less effort; but it requires intelligence and will power to avoid wasting time with detached words."[1]

[1] The Measurement of Intelligence, p. 274.

An alternative test for the tenth year in the Stanford scale is a simple form-board construction,[1] after Healy and Fernald. Four blocks are arranged in an irregular form before the child who is asked to arrange them into the frame so that they fit it exactly. The test is repeated three times within the space of five minutes, three successes being demanded. The examiner is interested in the time element and also in the method of procedure. The repetition of moves already found unsuccessful is a tendency of the dull. Terman places it as an alternative on the ground that its correlation with intelligence is lower than the majority of the tests used. It does not depend upon language performance and that is in its favour. But we shall give some attention to performance tests later, so need not go into detail at this juncture.

TWELVE YEARS.

The tests suggested by Binet for the twelve-year level of intelligence include the following: resisting suggestion, composing a sentence containing three given words, saying more than sixty words in three minutes, defining abstract terms, and reconstructing dissected sentences. Burt has an eleven-year-old test but all of the tests are given at earlier ages in the Stanford revision. For the twelfth year he has: giving three words to rhyme, rearranging dissected sentences, and the interpretation of pictures. The Stanford revision includes the vocabulary test, forty definitions with an equivalent of 7,200 words vocabulary being the standard for the age, definitions of abstractions, the superior plan of the ball-and-field test, rearrangement of dissected sentences, interpretation of fables, repetition of six digits reversed, interpretation of pictures, and giving the similarities of three things.

Binet's test of ability in resistence of suggestion has to do with length of lines which are shown successively to a child. Six pairs of lines each pair on a separate piece of paper are shown to the child. The first three pairs are lines of unequal length, the longer of the two being to the right, and each pair slightly longer than the one preceding. The three last pairs are of equal length. The child is shown each pair separately and is asked in the case of the first three which is the longer of the two lines. When the last three pairs are shown, the examiner asks each time: "And these?" The tests are passed if the child judges two of the last three pairs to be lines of equal length. Binet analyzes the test as one which brings two influences into play: (i) the influence of training, and (ii) the influence of reflection. The first three experiences have shown three unequal lines. The tendency is to suppose that this will continue. We have the beginnings of a habit forming

[1] For a discussion of the Form-Board tests, see pp. 87 ff.

process, an automatism. The second influence, reflection, has to resist the first in order to succeed. Success depends upon the careful perception of lines which will enable him to resist the suggestion formed by experience and tending to become automatic, and to perceive the lines as unequal. Suggestibility of this kind depends upon feelings and temperament as well as upon intelligence.

Terman follows Binet in making use of the test of definitions of abstract terms. Goddard, Kuhlmann and Bobertag also made use of the test, and there is fairly general agreement among them all as to the placing of the test, although Kuhlmann placed it in year eleven as did Binet himself in his early scale. Binet used the words *charity, justice* and *kindness.* Goddard followed him, translating *bonté* as *goodness* rather than *kindness.* Kuhlmann added *bravery* and *revenge.* Bobertag used *pity, envy,* and *justice.* Terman has *pity, revenge, charity, envy* and *justice.* Those who use three words demand two correct definitions out of three; those using five demand three correct ones out of the five. It need scarcely be pointed out that the ability to form abstract ideas calls upon the highest of the thought processes to function. It involves the processes of analysis and synthesis in which the properties of a number of concrete actions are analyzed and the common elements brought together in conceptual form. Obviously training would help the development of such ability, but intelligence would be a *sine qua non.* The mental defective is radically deficient in the power of generalization, so that the test at once marks him out. Even border line cases show marked inferiority in ability of this type. Of course there is some difficulty in the matter of interpreting definitions on the part of the examiner, but the instruction guides render the necessary help to the one who is beginning.

The rearrangement of dissected sentences is a test suggested to Binet by the "completion method" of Ebbinghaus. There is nowhere closer agreement about the placing of a test than in this case. Binet, Kuhlmann, Bobertag, Burt, Dougherty, Strong, Léviste and Morlé, Stanford University and Princeton University all agree in placing it here, Goddard alone holding it as an eleven-year test.

The following are the disarranged sentences which all use:

FOR THE STARTED AN WE COUNTRY EARLY AT HOUR
TO ASKED PAPER MY TEACHER CORRECT I MY
A DEFENDS DOG GOOD HIS BRAVELY MASTER

There are three possible solutions for the first, one for the second and two for the third sentence. One of each is:

We started for the country at an early hour.
I asked my teacher to correct my paper.
A good dog defends his master bravely.

The difference between the Ebbinghaus and the Binet method is that the former omitted parts of the sentence and required the subject to fill up the omissions whereas the Binet test gives all the parts and requires their arrangement in correct order. Says Terman : "The two experiments are psychologically similar in that they require the subject to relate given fragments into a meaningful whole. Success depends upon the ability of intelligence to utilize hints, or clues, and this in turn depends on the logical integrity of the associative processes. All but the highest grade of the feeble-minded fail with this test."

The Stanford revision introduces the fable-interpretation test. Five fables are used, viz., those of (a) Hercules and the Wagoner; (b) the Milkmaid and her Plans; (c) the Fox and the Crow; (d) the Farmer and the Stork; and (e) the Miller, his Son, and the Donkey. The following are the fables:

(*a*) [1]*Krishna and the Wagoner.*

"A man was driving along a country road, when the wheels suddenly sank in a deep rut. The man did nothing but look at the wagon and call loudly to Krishna to come and help him. Krishna came up, looked at the man, and said : 'Put your shoulder to the wheel, my man, and whip up your oxen.' Then he went away and left the driver."

(*b*) *The Milkmaid and her Plans.*

"A milkmaid was carrying her pail of milk on her head, and was thinking to herself thus : 'The money for this milk will buy 4 hens; the hens will lay at least 100 eggs; the eggs will produce at least 75 chicks; and with the money which the chicks will bring, I will buy a new dress to wear instead of the ragged one I have on.' At this moment she looked down at herself, trying to think how she would look in her new dress; but as she did so the pail of milk slipped from her head, and dashed upon the ground. Thus all her imaginary schemes perished in a moment."

(*c*) *The Fox and the Crow.*

"A crow, having stolen a bit of meat, perched on a tree and held it in her beak. A fox, seeing her, wished to secure the meat, and spoke to the crow thus : 'How handsome you are! and I have heard that the beauty of your voice is equal to that of your form and feathers. Will you not sing for me, so that I may judge whether this is true?' The crow was so pleased that she opened her mouth to sing and dropped the meat, which the fox immediately ate."

[1] The substitution of *Krishna* for *Hercules* is made for India.

(*d*) *The Farmer and the Stork.*

"A farmer set some traps to catch cranes which had been eating his seed. With them he caught a stork. The stork, which had not really been stealing, begged the farmer to spare his life, saying that he was a bird of excellent character, that he was not at all like the cranes, and that the farmer should have pity on him. But the farmer said: 'I have caught you with those robbers, and you will have to die with them'."

(*e*) *The Miller, his Son, and the Donkey.*

"A miller and his son were driving their donkey to a neighbouring town to sell him. They had not gone far when a child saw them and cried out: 'What fools those fellows are to be trudging along on foot, when one of them might be riding.' The old man, hearing this, made his son get on the donkey, while he himself walked. Soon, they came upon some men. 'Look,' said one of them, 'see that lazy boy riding while his old father has to walk.' On hearing this, the miller made his son get off, and climbed on the donkey himself. Further on they met a company of women, who shouted out: 'Why, you lazy old fellow, to ride along so comfortably while your poor boy there can hardly keep pace by the side of you!' And the poor good-natured miller took his son up behind him, and both of them rode. As they came to the town a citizen said to them, 'Why, you cruel fellows! You two are better able to carry the poor little donkey than he is to carry you.' 'Very well,' said the miller, 'we will try.' So both of them jumped to the ground, got some ropes, tied the donkey's legs to a pole and tried to carry him. But as they crossed the bridge the donkey became frightened, kicked loose, and fell into the stream."

After reading a fable to the child he is then asked to tell what lesson it teaches us. The response is scored as correct when the pupil interprets the fable correctly in general terms, and is given a half score when the interpretation is in general terms and fairly plausible though not accurate, or when it is substantially correct though not generalized. Terman says that the test may aptly be called the test of the power of generalization. Its psychological value is that it is analogical of many situations which occur in actual experience, calling for an exercise of responses to social stimuli. This is at the basis of all ethical behaviour, and gives us a clue to the reason that a mentally defective person is unable to be moral. It is not the case of being radically opposed to existing conventions or traditions, that leads the feeble-minded person to show apparent disrespect for received standards and customs. The reason is that he has not the intelligence to generalize so as to understand that a certain situation belongs to a certain class of

situations demanding a certain type of response on his part. Moral judgements are social judgements, and investigation shows that many of the criminal and delinquent classes are immoral because they are unsocial, and they are unsocial for lack of intelligence. Hence a test which measures a child's ability to generalize is of inestimable value in determining the place which he is capable of occupying in the social order. It presents an imaginary problem which if he is able to solve indicates his ability to meet a moral situation when faced with it, and if he is unable to solve indicates the reverse.

The other tests employed do not involve any new psychological elements which we need to consider. They call for the same types of responses as those already considered, the difference being simply a matter of complexity, it being understood that the mental processes develop in their ability to meet complex situations with advancing years.

The only other notable scale besides the Binet and the revisions of it which we have considered is the Yerkes Point Scale. And, as already indicated, the fundamental difference is not one of type but rather of method of scoring. So that it will not be necessary to discuss the tests as we have already dealt with all the types and with the majority of the actual tests used in the lower grades. The Yerkes Point Scale is a single scale and not one divided into sections corresponding to age. There are twenty tests, as follows: æsthetic discrimination, indicating omissions from pictures, discrimination of lines and weights, memory span for digits, counting in inverse order, repetition of words and sentences, reaction to pictures (whether enumeration, description, or interpretation), arrangement of weights, comparison of concrete objects from memory, definition of concrete objects in terms of use, resistance to suggestion, copying figures, giving number of words in three minutes, writing sentences containing three given words, comprehension tests, drawing designs from memory, criticism of absurd statements, reconstructing dissected sentences, definitions of abstractions, and completing analogies. Each child is tested on the whole performance and each test is given a numerical scoring value. Then the total score gives the value of the child's intelligence in terms of the Point Scale. The scores have been equated with mental ages, and a complete table may be consulted in Yoakum and Yerkes: *Army Mental Tests*, p. 97. I quote a few as illustrative:

Score.	Mental age.	Score.	Mental age.
88 to 100	18 or above	60	10·3
87	17·5	50	9
86	17	40	7·8
80	14·5	39	7·7 or above
70	12	38	7·5

Score.	Mental Age.	Score.	Mental Age.
37	7·3	30	6·3
36	7·2	20	4·7
35	7	15	4

A perusal of the comparisons will show what one would expect, viz., that it is possible to determine in terms of much finer measurements the exact mentality of the subject than it is higher up in the scale. The difference of one in a score makes a difference of half a year in mental age when one is at the top of the scale, whereas it makes a difference of only one-tenth or one-fifth of a year at the lower end. This is the mechanics of the fact that mental processes are more simple and therefore more readily measurable in young children, and more complex and hence more difficult of exact appraisal in adults.

The other intelligence tests for young children that are in use are group tests, and will therefore fall to be discussed in the lecture dealing with them. Most investigators do not give up the individual tests when they undertake the group tests, but use the two together. A study of the correlation of the results of the two is also valuable.

CHAPTER IV.

INTELLIGENCE TESTS FOR SENIOR GRADES.

We may speak of the junior grades as occupying the period known as childhood, and the senior grades as adolescence. We are therefore concerned in this chapter with tests that measure the intelligence of adolescents. We do well to observe that we are concerned with a period of life that is psychologically quite distinct from the previous and succeeding periods. Broadly speaking, it may be delimited as the period which begins with the dawning of the sexual life or puberty and ends with maturity. In actual life there is a good deal of variation in the beginning and ending of the period, but the period both begins and terminates a little earlier in females than in males. Physiologically speaking, the period begins about two years later than it does psychologically. It is a period of marked changes in the child, and these changes begin to be apparent in the mental life a year or two earlier than they are in the physical life.

It is not necessary for our purposes to go into the matter of the subdivisions of the adolescent period which psychologists have observed. Suffice it to note that it is by no means a static period, but that it is marked by a process of unfolding mentally as well as physically. The adolescent period is marked by the birth of a larger self. There is a desire for a larger realization of the self through self-assertion and self-help, due to the fact that new forces are beginning to operate, and new powers to function. This expresses itself in the reaching out socially as well as in an increased sense of individuality. At the same time, it is a period characterized by contradictions and anomalies. One can never be quite sure what to expect from the adolescent youth. The rapid physical growth, which is accompanied by the beginning to function of higher intellectual powers and an enlarged social consciousness, means that the child is being born into a new world, larger and at the outset full of bewilderment. Professor Stanley Hall has characterized the period as one of "alterations between excitement and inertness, pleasure and pain, self-confidence and humility, selfishness and altruism, society and solitude, sensativeness and dullness, knowing and doing, conservatism and iconoclasm, sense and intellect, wisdom and folly."

The adolescent period is a period of new intellectual alertness. It is a period in which the thinking processes are suddenly and vigorously stimulated into greater activity. It comes out in the tendency to ask questions about many things which before have been accepted on faith. There is a much broader range of interest than hitherto, an evidence of the expansion of conative functions. The instinct of curiosity begins to be much more active, so that the

child is much more inclined to investigate and explore into new avenues of life. He is not so contented with the authority of his elders. These are facts which are of value to the educator who must understand the psychological characteristics of the period, if he is to deal with it intelligently. Psychological tests bear out the truth of these remarks in disclosing an ability to undertake mental tasks which call for an increasing power of exploration, and should be so designed as to test that development.

The adolescent years have been described as "the socializing years". There is a demand for a larger social life than the family is able to satisfy. It is the friendship forming period. The educator is particularly concerned with this fact, because of the fact that the social environment has an important function to play in the development of personality. The ability to respond to the demands of life is to some extent in proportion to the helpfulness or otherwise of the social environment, and is reflected in tests of mental abilities.

Adolescence is marked by the consciousness of high aspiration. Here we see the youth's admirations for the attainments of maturity. It is more or less the period of hero-worship. This tendency takes the form of emulative and imitative activities. The innate tendency to imitate is developed and is tied up to the idealism of the hero-worshipper. The ethical significance of this fact is obviously very large. The psychological phenomenon itself is one which marks the growing intellectual alertness, and reveals itself in practical tests to which the youth is put.

The adolescent period is a period of stress and strain. This expresses itself often in friction against one's surroundings, and constant endeavours to do new things, see new places, and know new people. This storm and stress does not always take the same form of expression. Sometimes it makes for morbid introspection, brooding and depression; sometimes for hilariousness and uncontrolled spirits; sometimes for abnormal self-consciousness and bashfulness. The educator who would test the intelligence of an adolescent youth must bear this in mind, and be sure that the conditions under which the test is given are not such as to invalidate the results because of any of these phenomena. The life processes are welling up into a larger life, and, if tactfully directed, will attain their maximum of development.

Educational methods must take wise cognizance of these facts in regard to the psychological characteristics of the period. The clearer the knowledge of the natural tendencies and dispositions, the better will the educationalist be able to minister to the youth. It becomes his duty to bring the natural tendencies to a successful issue without dwarfing the self that is developing towards maturity. Outlets for the newly aroused activity must be afforded. Abundant opportunities for satisfying his expanding sense of

selfhood in wholesome channels must be presented. The youth's presuppositions must be captured in favour of the higher standards of value. His habits must be formed in such a way as to prove permanently serviceful both to himself and to others. The tendency to self-assertiveness must be directed to the attainment of self-mastery in this time of new adjustments. These are phases of development which the psychologist wants to watch, and standardized measurements ought to be so designed as to enable him to judge to what extent the desiderated expansion is being procured. They ought to be diagnostic of any ills that need attention, and they ought to be devised so as to appeal to these expanding abilities.

The tests for measuring the mental abilities of adolescents and adults which were originally devised by Binet underwent many modifications. Even at his own hands there were a number of changes. One of the problem tests, for example, was placed by him in the twelfth year in his 1908 scale, but advanced to the fifteenth year in the 1911 revision. The same is true of the test of repeating seven digits. The problem of reversing the hands of a clock Binet placed in his 1905 series, but omitted from his subsequent revisions. Burt has followed Binet in having tests for the years thirteen, fourteen and fifteen, but strangely he has only two tests for each of the years thirteen and fourteen, and five for the fifteenth year. At the same time, he confesses that the tests are quite inadequate and sets out on a new line by the substitution of reasoning tests for the revised Binet tests. Terman radically departs from the Binet tests, though he uses several of them. He has tests for the fourteenth year, for the average adult, and for the superior adult, and his tests have proven to be about the most satisfactory as individual tests of adolescent and adults. We shall consider them in the order in which he recommends them.

FOURTEEN YEARS.

The first test for this year in the Stanford revision is the vocabulary test. The same list of words which was used in the eighth, tenth and twelfth years is used. A fourteen-year-old child should be able to give fifty correct definitions which at the calculation made by these investigators indicates a vocabulary of approximately 9,000 words.

The next test is called by Terman the *Induction Test*: *finding a rule.* The experimenter provides six sheets of thin blank paper, say 8½ × 11 inches. The first sheet is folded before the child, and a small piece cut or torn out of the folded side. The child is asked to tell how many holes there will be in the paper when unfolded. The correct answer is usually forthcoming with no difficulty. Whether it be right or not, the experimenter unfolds the paper, and exhibits it for the inspection of the subject. Then

he repeats the experiment with the second paper, folding it twice, again exhibiting it after securing the subject's response. This is repeated for the six sheets, in each case recapitulating the results before proceeding to the next experiment. The tests are scored as successfully passed if the child realizes the rule by the time that the sixth sheet is reached, even though he makes five incorrect responses, providing the sixth be correct and the child discovers the rule by this inductive process. No hint should be given of course that there is any rule by which the matter can be determined, but the child is left free to discover it for himself. The test is well named the *Induction Test* for it is by the logical process of inducing from particulars to general that the child is able to discover that there is a rule operating whereby one can foretell what will happen in the next case. Very few people, even adults, have been found to reason it on a deductive basis. The Stanford investigators have found that it is a test of intelligence which is influenced to a minimum degree by schooling, and that it has the added advantage of being free from language difficulties. The ability tested is that of generalizing from particulars, a process of abstraction, and the fact that experiments indicate that it is almost invariably arrived at by a process of induction shows that it has called into exercise processes of exploration for which the adolescent is noted. The test seldom fails to arouse interest and attention, so that the child enters heartily into the attainment of a solution, especially if it be presented to him as in the form of a puzzle.

For age fifteen Burt makes use of the same type of test, much modified. He suggests only two sheets of paper, one of which is folded in four like an envelope, and in the middle of the edge which presents but a single fold a triangular notch about one cm. deep is drawn. Then the instructor says to the child : " Here is a sheet of paper that has been folded across, and then folded again. Now suppose I cut a notch just here. When the paper is unfolded again, what would it look like ? Will you show me on this piece how and where it would be cut ? " The child is scored as having responded correctly if he draws two diamond-shaped holes in a line with each other, each in the middle of one-half of the paper. It is apparent that this test is at once more difficult and easier than the form in which it is given by the Stanford group. Moreover, the Stanford form of the test calls forth the ability of the child to induce and abstract a general rule on the basis of observed particulars which the Burt form of the test is not so well fitted to do because of insufficient particulars. At the same time, the Burt form of the test calls for deeper thought and imagination for the same reason that particulars are few.

Binet is responsible for the test of giving differences between a president and a king. Many of the revisers omit it, especially those in countries where there is no president. It has been suggested that in India the test be in the form of the difference between a governor

and a viceroy. Burt remarks upon it as a test that is obviously better suited to French and American than to British children. Still the kingship is as strange to many American and French children as is the presidency to British children. Terman places the test in those for year fourteen asking the child to state three main differences between the two offices. He states that, were only one difference required, the test would be suitable for the twelfth year. The three differences which are expected relate to manner of accession, tenure of office, and degree of power. Sometimes children state differences that are trivial and insignificant, but these are not scored as successes. It is only when one of the chief differences is stated that the child is given credit. Terman's experience is that about 30 per cent of "average adults," including high school students, will state at least one unsatisfactory contrast. Some criticism has been levelled against the test as demanding too much schooling, and this would be true if it were applied to children that were very young. But it may be defended as a test of intelligence of the fourteen-year level, as Terman has indicated, on the ground that at such a developed stage it tests the power of discrimination, and that it would be difficult to find a person of that age of mentality, no matter how poor had been his educational advantages, who could not respond correctly. Even some who are feeble-minded are able to answer, their difficulty being not lack of knowledge of the facts, but possession also of a number of irrelevant or trivial facts, and inability to discriminate the principal from the unimportant distinctions between the two offices. The psychological features of the test correspond to such earlier tests as the stating of similarities and differences which call into play the associative tendencies. In this test, however, we have the added factor of discrimination between the important and the relatively insignificant.

Another test is that of the problem question, which the Stanford revisers placed in the fourteenth year, after a very extensive testing of the test. Binet had placed it in the twelfth year in his 1908 scale, and had put it on to the fifteenth year in his 1911 revision. Goddard and Kuhlmann retained it as a twelve-year test. The child's attention is secured whereupon he is asked to give such an answer to each problem as will show that he has understood it. Of the three problems given, Binet constructed the first two and Terman the third.

(*a*) "A man who was walking in the woods near a city stopped suddenly, very much frightened, and then ran to the nearest policeman, saying that he had just seen hanging from the limb of a tree a . . . a what?"

(*b*) "My neighbour has been having queer visitors. First a doctor came to the house, then a lawyer, then a minister (priest, clergyman, or preacher). What do you think happened there?"

(*c*) "An Indian who had come to town for the first time in his life saw a white man riding along the street. As the white man rode by, the Indian said—'The white man is lazy; he walks sitting down.' What was the white man riding on that caused the Indian to say, 'He walks sitting down'?"

The test is one form of the completion test which we have noticed before. Some of the elements of the situation are given on the basis of which the subject is expected to reconstruct the entire situation. As pointed out before, this type of test calls for a certain amount of exploring among the associations that lie dormant in order to find the appropriate one. Success depends upon the ability to make use of hints and clues which ultimately depends upon the integrity of the associative processes. It need scarcely be added that the correct solutions to the problems are (*a*) a corpse, (*b*) a death, and (*c*) a bicycle.

Terman introduces into the fourteen-year series an arithmetical reasoning test, consisting of problems selected from Bonser in *Columbia University Contribution to Education.* The problems may be adapted to India as follows:

(*a*) If a man's salary is Rs. 20 a week and he spends Rs. 14 a week, how long will it take him to save Rs. 300?

(*b*) If two fountain pens cost Rs. 5, how many pens can you buy for Rs. 50?

(*c*) At as. 6 a yard, how much will 7 feet of cloth cost?

It has sometimes been objected that these problems depend not so much upon intelligence as upon schooling. To be sure, the subject undoubtedly makes use of knowledge which he has acquired in school, that is of the knowledge of the way of working the elementary arithmetical processes. But a successful manipulation of these elementary processes themselves involves intelligence. Terman says: "Success depends upon the ability to apply this knowledge readily and accurately to the problems given—precisely the kind of ability in which a deficiency cannot be made good by school training. We can teach even morons how to read problems and how to add, subtract, multiply, and divide with a fair degree of accuracy; the trouble comes when they try to decide which of these processes the problem calls for. This may require intelligence of high or low order, according to the difficulty of the problem."[1]

Reversing the hands of a clock is a good test of constructive visual imagery. The test is conducted by the experimenter saying to the subject: "Suppose it is six-twenty-two o'clock, that is twenty-two minutes after six; can you say in your mind where the large hand would be, and where the small hand would be?" After securing assent, he continues: "Now, suppose the two

---

[1] The Measurement of Intelligence, p. 320.

hands of the clock were to trade places, so that the large hand takes the place where the small hand was, and the small hand takes the place where the large hand was. What time would it then be?" The test is repeated for three different times of day, namely, 6'22, 8'10 and 2'46. The correct answer to the first falls between 4'30 and 4'35, the second between 1'40 and 1'45, and the third between 9'10 and 9'15. The subject is not permitted to look at any time-piece or to help himself by means of a drawing, but must work the problem mentally. This test illustrates very well a point that is much discussed by psychologists, namely, whether or not thinking involves imagery, and if so what types of images prevail. This test, as already indicated, obviously depends on the ability of the subject to visualize and to control his constructive visual imagery. There have been instances however where correct solutions have been attained on the basis of verbal imagery employed in a strictly mathematical process. Subjects who are not accustomed to employ much visual imagery, however, as a rule find great difficulty in solving this type of problem. The fact that the majority of those of fourteen-year intelligence are able to solve the problem argues strongly that the thinking of most people is in terms of visual imagery. The manipulation of imagery depends partly upon the vividness of the original sense images. The recalled image is usually fainter than the original, and the fainter the imagery the more difficult it is for the person to solve problems which involve constructiveness. Whipple in his *Manual*[1] has given a number of tests which measure the imaginative processes, and shows that there is a high positive correlation between success in such tests and intelligence.

AVERAGE ADULT.

The tests for the average adult in the Stanford revision include the vocabulary test, the interpretation of fables (higher score), the giving of differences between abstract terms, the problem of the enclosed boxes, the repetition of six digits reversed, using a code, and two alternative tests of repeating twenty-eight syllables and of comprehending physical relations. In the vocabulary test the average adult is expected to be able to give 65 out of the 100 definitions which at the calculation used indicates a vocabulary of 11,700 words. The fable interpretation test is conducted as in the twelfth year, except that the standard demanded is higher. The interpretation of the twelve-year-old was expected to include five points; that of the average adult should include eight points.

The differentiation of meaning between pairs of abstract terms was devised by Binet and first used in his 1908 scale as a test of thirteen-year-old intelligence. He suggested five pairs of words to

[1] Vol. II, pp. 619–673.

be differentiated, the bracketed words being those originally used in English—

(i) paresse, oisiveté (poverty, misery);
(ii) évenement, avenement (event, advent);
(iii) evolution, revolution (evolution, revolution);
(iv) plaisir, bonheur (happiness, honour);
(v) orgeuil, pretention (pride, pretence).

In his 1911 revision the last two pairs were dropped; and the other three pairs were moved to the adult group. Terman dropped also the event-advent pair and added two new ones, namely laziness-idleness and character-reputation. Three correct attempts out of four are required for a pass. Naturally there is considerable variety possible in the correct answers which may be given, but by practice the experimenter will be able to discriminate and grade the responses. The test calls for the same type of psychological processes as the twelve-year test where abstract terms are defined, but is of greater difficulty in that a comparison has to be made. It involves processes of abstraction which mark the advancing intelligence of an adult, and could not be expected of an undeveloped intelligence. At the same time success depends upon the power of expression to such a considerable degree that it would not do at all for a test in any case where the language difficulty appeared.

The problem of the enclosed boxes is put to the child by showing him a small cardboard box, and saying to him: "You see this box; it has two smaller boxes inside of it, and each of the smaller boxes contains a little tiny box. How many boxes are there altogether, counting the big one?" After recording the response the test is repeated with the difference that the subject is told that each of the smaller boxes contains two tiny ones. A third time it is varied so that there are three smaller boxes, each containing three tiny ones. The fourth time there are four smaller boxes each containing four tiny ones. The problem is given and solved orally, three correct solutions out of four being scored as a success. Here again constructive imagination is called into function, and success waits upon the ability to manipulate concrete visual imagery. At the same time it resembles the problem of reversing the hands of a clock in that it is solved by some subjects by means of verbal imagery in a mathematical process. Imagery of the tactual type would probably serve with some persons.

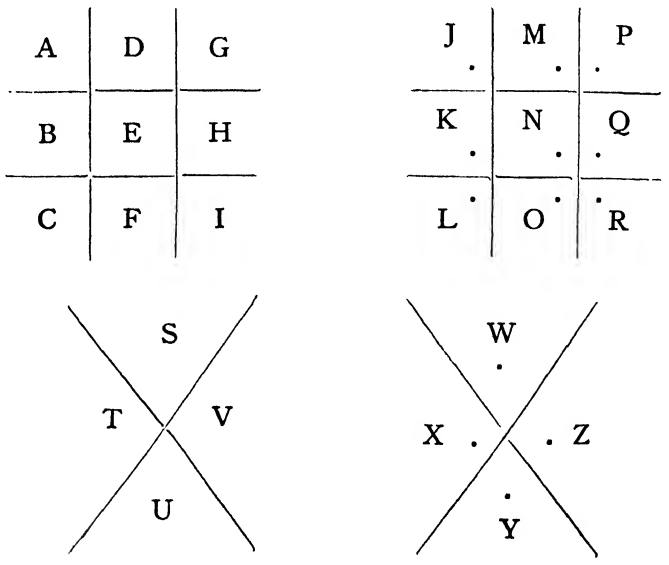
Terman remarks in commenting on this particular test that "this is as good a place as any to emphasize the fact that the introspective study of mental imagery has little to contribute to the measurement of intelligence. Intelligence tests are concerned with the total result of a thought process, rather than with the imagery supports of that process. Thoughts may be carried on almost equally well by various kinds of imagery . . . We may say that imagery is to thinking what scaffolding is to architecture. The

important thing is the completed building rather than the nature of the scaffolding employed in erecting it. No one thinks of blaming the ill-construction of a building upon the scaffolding used, for if the architect and builder are competent, satisfactory scaffolding will be found. Just as little are deficiencies or peculiarities of imagery the real cause of low-order intelligence. We cannot increase intelligence by formal drill in the use of the supposedly important kinds of mental imagery, any more than we can transform a plain carpenter into a Michael Angelo by instructing him in the use of scaffolding materials such as were employed in the construction of St. Paul's Cathedral."[1] It seems to me that Terman has combined fact and fiction in these comments into rather unsound conclusions. While it may be true that the introspective study of imaginative processes does not supply us with a criterion for the measurement of intelligence, it is also true that the measurement of intelligence by means of tests, such as this problem of the boxes and the other of telling the time were the hands of the clock reversed, throws considerable light on the manner and significance of image manipulation. And that in turn has value for us in suggesting elements which we must not neglect in the devising of tests. Again, while it may be true that thought may be carried on by various kinds of imagery, it is doubtful whether Terman is justified in using the qualifying phrase, "equally well." The fact of the matter is that his own investigations show that people who are deficient in visual imagery largely fail in such tests because the majority of people do the major portion of their thinking by means of visual imagery, and surely here, if anywhere, we are concerned with what *is* rather than with what *may be*. It is very doubtful whether most people use visual imagery in preference to other types of imagery as an accident. Doubtless our habits contribute, but that "innate preference" whereby we select has no doubt taught us to employ visual imagery as the most serviceable and economical in reflective experience. Further I dissent from Terman's description of the relation of imagery to thinking by the analogy of a scaffolding's relation to architecture. If mental imagery is only the scaffolding I shoud like to know with what materials the learned Professor would propose to construct the buildings of thought. It is surely much truer to liken imagery to the very building materials themselves. Images are the stuff of our thinking. One can no more think without images of any kind, than he can erect a building without bricks and mortar or other building materials. And if this analogy be truer to the facts, it means that a greater amount of stress should be placed on the significance of imagery and its manipulation than Terman

[1] The Measurement of Intelligence, pp. 328–329.

suggests. He has been quoted as maintaining that formal drill in the use of various kinds of images will not increase intelligence. But one's understanding of his earlier treatment of the subject would be that he regards intelligence as congenital having reference to one's native ability in contrast to his acquirements. If that be the case, it is doubtful whether any kind of drill would actually increase intelligence, though on the other hand everybody would admit that intelligence would thereby be trained for greater service. An intelligence that is supplemented by real attainments will be of greater individual and social worth than one that is raw and untrained, other things being equal. Drill in the construction of scaffolding would not account for the difference between a plain carpenter and Michael Angelo, but training in the judicious use of building materials would have more far-reaching effects in architecture than attention only to scaffolding construction. All that we can do to develop the child's ability to observe and to retain his observations as images, subject to recall when needed, will be of immense service educationally, and by that I mean observation in the larger sense of sense-perception. The more he attends to the collection of materials, the better outlook for a good building.

The test of using a code was one that was devised by Healy and Fernald, and was described in their *Tests for Practical Mental Classification*. Goddard made use of it as a test for fifteen-year mentality in his revision of the Binet scale, and the Stanford revisers placed it as a test of average adult intelligence which they equated with sixteen years. The subject is shown the code as given in the following form:—



Then he is asked to look carefully and note the arrangement of the letters. He will be directed to the facts that the first two diagrams have the letters in the up-and-down order, whereas the third and fourth are arranged in reverse order to the hands of a clock. The second and the fourth resemble the first and the third respectively, except that they have dots in each corner. Then he is told that this represents a code, not a play-code but a real code which was actually used in sending communications in the American Civil War. The secret messages were sent by drawing the lines which hold a letter, including the dots where necessary, but without the letters. The subject is then shown how to use the code by the use of an illustration or two, such as the words *war* and *spy*: after this illustration of the use of the code, the diagrams are removed and the subject is asked to write the words "COME QUICKLY" in code form, without reproducing the entire code on paper. The test is scored a success if the subject writes the two words within six minutes with not more than two errors, the omission of a dot counting as one-half mistake. Healy and Fernald, who originated the test, described it as one which measured "close attention and steadiness of purpose." They also mention that the attention must have an inward direction since there is no external object to which the sense-organs can refer for stimulus and help. Terman relates that, contrary to their expectations, the use of visual imagery was not particularly necessary to the result, but that kinæsthetic imagery would serve the purpose equally well. Auditory-verbal imagery would also serve the purpose. He has also ascertained that nearly all subjects over twelve-year intelligence who fail on the test are nevertheless able to reproduce the diagrams and insert the letters in their correct spaces. This seems to indicate that the actual use of the code demands a much more focalized type of attention than does the mere remembering of the code itself. Terman also observed that "high school pupils for some reason not apparent" were more successful in the test than were unschooled adults of the same mental ability. Perhaps the solution is to be found in the fact that a trained intelligence of a certain inherent capacity will make certain responses better than an untrained intelligence of the same capacity, because the training, though it may have been directed to a different set of responses, has called into function elements of intelligence that are fundamental to the test under observation. One could quite conceive of this test being useful in the vocational selection of operators for the telegraph department.

As an alternative test, I have already observed that the Stanford revision includes the repetition of twenty-eight syllables. The sentences used in the test are as follows:—

(i) "Walter likes very much to go on visits to his grandmother, because she always tells him many funny stories."

(ii) " Yesterday I saw a pretty little dog in the street. It had curly brown hair, short legs, and a long tail."

The test is scored as a success if the subject repeats one of the two without a single error. This type of test is not as satisfactory for the higher levels of mentality as for the junior grades, as it is "too mechanical to tax heavily the higher thought processes." This test has appeared several times before, and it might be well at this point to quote Burt's calculation as to syllable repetition in relation to mental age. It is as follows:—

age four	...	6 syllables;
age five	...	10 syllables;
age seven	...	16 syllables;
age fourteen	...	26 syllables.

A second alternative test in the Stanford revision for the average adult is in the form of problems involving the comprehension of physical relations. Three problems are given out of which two correct responses are required to score success. These are—

(i) problem regarding the path of a cannon ball;
(ii) problem as to the weight of a fish in water;
(iii) problem of the difficulty in hitting a distant target.

In the first problem there is drawn on a piece of paper two parallel horizontal lines one of which is about eight inches and the other about one inch long. The first represents the level ground of a field, and the second a cannon, pointed horizontally, parallel with the level of the ground. The subject is then told: "Now, suppose that this cannon is fired off and that the ball comes to the ground at this point (pointing to the farther end of the line which represents the field). Take this pencil and draw a line which will show what path the cannon ball will take from the time it leaves the mouth of the cannon till it strikes the ground." The only correct answer is that which describes the path of the cannon ball as almost on a level at the beginning and then as dropping more rapidly towards the end of the course. The second problem is: " You know, of course, that water holds up a fish that is placed in it. Well, here is a problem. Suppose we have a bucket which is partly full of water. We place the bucket on the scales and find that with the water in it it weighs exactly 45 pounds. Then we put a five-pound fish into the bucket of water. Now, what will the whole thing weigh?" Many will answer 50 pounds at once, but when they are asked how that can be, since the water itself holds up the fish, will apologize for answering thoughtlessly. The answer is only scored correct when the subject adheres to that answer on the ground that the scales have to hold up the total weight of bucket, water and fish. Problem three is stated thus: " You know, do you not, what it means when they say a gun 'carries 100 yards?' It means that the bullet goes that far before the bullet

drops to amount to anything. Now, suppose a man is shooting at a mark about the size of a quart can. His rifle carries perfectly more than 100 yards. With such a gun, is it any harder to hit the mark at 100 yards than it is at 50 yards?" After the subject responds, he is asked to give reasons for his answer. The only correct answer is one which shows that the subject appreciates the fact that a deviation from the mark due to incorrect aim would become wider at 100 yards than at 50 yards. Terman, who devised this test, defends it very properly on the ground that the ordinary experiences of life lead one to comprehend the commoner physical relationships, even when the subject has not had the opportunity of schooling. Success depends on the innate tendency to explore the unknown, and to pry into the secrets of natural phenomena. Many times the observations will be quite correctly formed where the subject has not learned the underlying reasons. It is perfectly legitimate to standardize these products of the natural observational tendencies as indicative of the development of intelligence. Terman gives a long list of the commoner physical relationships, a list which might be much expanded, of observations that it would be possible by experimentation to standardize in respect to the mental levels which they indicate. Such phenomena might be included as that an unsupported object falls to the ground, that fire burns, that birds fly in the air, that water will not run uphill, that it is hard to run against a strong wind, that a heavy object is harder to move than a light one, that sounds are sometimes followed by echoes, that the heart beats faster and the rate of breathing is increased by running, and so on *ad libatum*.

SUPERIOR ADULT.

The Terman tests for the superior adult are as follows: the vocabulary test, Binet's paper-cutting test, the repetition of eight digits, giving the thought of a passage, the repetition of seven digits reversed, and what he calls the ingenuity test. The digit-repeating test both in regular and reverse order has been discussed, the only difference here being the increased difficulty due to the greater number of digits to be remembered. The vocabulary test for the superior adult is standardized for seventy-five definitions which is calculated to indicate a vocabulary of 13,500 words.

The paper-cutting test is another application of the same problem which appeared in the induction test of year fourteen. In this instance the experimenter takes the piece of paper, and asks the subject to watch as he folds it at right angles twice across the middle, and then cuts a notch in the middle of the side presenting one edge. Then the subject is given a second piece of paper like the first and asked to make a drawing to show how the first piece of paper would appear if it were unfolded, by drawing lines representing the creases and making marks to indicate the results of

the cutting. The test is scored correct when the creases are drawn correctly and the holes are located properly, irrespective of the shape of the holes. Here again we have a test which depends for its success upon the correct manipulation of visual imagery. It is not enough to be able to carry the images in a memory process, but there must be ability in constructively combining them. This is a test that does not depend upon educational advantages for in many cases the unschooled subjects succeed better than the schooled. Terman also states that "it appears that a solution is seldom arrived at, even in the case of college students, by logical mathematical thinking."[1]

The test of repeating the thought of a passage is one which was devised by Binet, and serves as a comprehension test rather than as a pure memory test, as one might suspect. Before the passage is read the person is asked to attend with the object of afterwards giving in his own words the substance of the passage read. Two selections are used, as follows:—

(i) "Tests such as we are now making are of value both for the advancement of science and for the information of the person who is tested. It is important for science to learn how people differ and on what factors these differences depend. If we can separate the influence of heredity from the influence of environment, we may be able to apply our knowledge so as to guide human development. We may thus in some cases correct defects and develop abilities which we might otherwise neglect."

(ii) "Many opinions have been given on the value of life. Some call it good, others call it bad. It would be nearer correct to say that it is mediocre; for on the one hand, our happiness is never as great as we should like, and on the other hand, our misfortunes are never so great as our enemies would wish for us. It is this mediocrity of life which prevents us from being radically unjust."

The test is scored as a success if the subject can repeat in fairly consecutive order the principal thoughts in either of the passages read, no attention being given either to style or *verbatim* repetition. In other words, it is employed purely as a test of thorough comprehension. This is another of that type of tests where a great variety of responses is obtained with varying degrees of accuracy. It can be only by practice and care that the experimenter learns which responses to score as correct and which as unsatisfactory. The difficulty inherent in these problems is that they deal with abstract matters, and the mentally deficient cannot do very much with abstractions, their thinking clinging, as Terman says, "tenaciously to the concrete." This type of test calls for conceptual analysis and synthesis in which the contents of concrete experiences are broken up into relatively elementary factors which are again

[1] The Measurement of Intelligence, p. 339.

recombined into new mental constructs. Ideational activity is differentiable from perceptual precisely on this basis that it involves generalization to some degree. There is nothing to hinder even the mentally defective who has a normal set of sense organs and a healthy nervous system from carrying on the processes of sense-perception which are involved in the attainment of concrete knowledge. But the conceptual process calls for processes of analysis and synthesis which demand abstract thinking of which the mental defective is constitutionally incapable. From the point of view of the psychological processes involved the test is quite legitimate. The only difficulties involved are those of language and of dependence on schooling which make it rather unsatisfactory for a few subjects who are really "superior adults."

The ingenuity test consists of three similar problems. The first is stated as follows:—

"A mother sent her boy to the river and told him to bring back exactly seven pints of water. She gave him a three-pint vessel and a five-pint vessel. Show me how the boy can measure out exactly seven pints of water, using nothing but these two vessels and not guessing at the amount. You should begin by filling the five-pint vessel first. Remember, you have a five-pint vessel and a three-pint vessel, and you must bring back exactly seven pints."

The second problem resembles the first except that the subject is to bring eight pints with a five-pint and a seven-pint vessel, beginning by filling the five-pint one. In the third problem seven pints are to be brought with four and nine pint vessels, beginning with the four-pint vessel. A time limit of five minutes is set for each problem, and two correct solutions out of three are scored as a success. The problems are stated orally, are worked without the assistance of pencil and paper, and the solution must be presented orally as a complete record of the method to be used. This test was devised by Terman when making a study of the mental processes of bright and dull boys, but experimentation with it led him to see that it demanded a much higher degree of mentality, so that eventually it was standardized as a test of "superior adult" intelligence. In the main, success depends upon the functioning of what we might call the creative element in intelligence which is involved in practical judgement and in invention. It calls into operation similar processes to those which are employed in the creative imagination of the scientific worker. This ability accounts for the fact that cultured man uses a spade and a fork where primitive man used a grubbing-stick, that he lives in houses where his uncivilized ancestor lived in caves, and so on. Psychologically speaking, ability to solve such tests as this depends, as do inventive operations generally, upon the ability to analyze, abstract, manipulate imagery, and adapt the conceptual results to new situations.

The creative tasks in life are not accomplished by the person who can think only perceptually. But we owe much of our progress in art, in science, in religion, and in philosophy to the few men of superior intelligence who bring to life's problems the ability to analyze and to synthesize in new and untried ways. After all the method of trial-and-error is responsible in actual life for much of our advance. But it takes a man of unusual ability as a conceptual thinker to make such abstractions and devise such new syntheses as make progress possible. The originality and individuality of the genius account for many of the inventions that have proved of the largest service to the human race. If we are therefore able to discover by psychological tests the presence of superior intelligence, the social possibilities of developing it to its utmost capabilities are greatly enhanced. The task of discovering the superior intelligents is equally as important as that of selecting the inferiors. If the latter are a danger to the community, the former are its latent power. Yet experience has shown that very often the superior person is less likely to be discovered than the inferior, so that much latent power is left unharnessed.

School teachers who have no other technique than the examination method, by which to classify their pupils, very often fail to detect the children of superior intelligence. They may be described as "doing good work," or sometimes as "fair but showing no unusual ability." And when the intelligence test is introduced it is found that the child is capable of doing much more advanced work than that which is being given. The work of the class is making no demand on the intelligence of the child, and failing to call out any constructive ability. The work of the class may be so much behind the child's ability that it fails to elicit any real interest without which normal development cannot take place. Prof. Whipple of the University of Illinois has interested himself in this problem, and has been conducting an experiment with children of superior intelligence. The aim of the experiment was to ascertain how much progress was possible if a class of all superior intelligents were put together and allowed to work, without crowding, as much as they were capable. Care was taken in the experiment to see that there was nothing unusual or distinctive about the room except the superior intelligence of the students. Thirty pupils were selected, fifteen of the fifth grade and fifteen of the sixth grade, all of them superiors. They were under the instruction of a well trained teacher and their progress was observed by means of educational and psychological tests throughout the year. With no impairment of health to the pupils they were enabled to cover in one year as much as is covered in the curriculum for two years' work. There was practically no occasion for discipline, attendance was above the average for other classes, and there was no evidence of self-conceit or clannishness. If this experiment can be accepted as

typical of what may be done anywhere under ordinary conditions, it is symptomatic of a waste of time in the case of many brighter pupils and a consequent neglect of conditions under which the best development can be secured.

The intelligence test, because it enables the educationalist to classify his pupils on a more scientific basis, thus secures justice not only for the inferiors and the superiors, but also for the average child. School organization cannot be thoroughly scientific unless it takes account of mental capacity, and the test is a device that will enable us to obtain that exactness which ought to characterize any discipline that claims to be a science.

CHAPTER V.

PERFORMANCE TESTS.

Reference was made in the first chapter to the beginnings of the performance tests. It was observed that the Binet type of test was open to the criticism that it demanded a comprehension of language and also an adequate language response. Obviously a test or scale of tests that rested so much on a language basis would not be of service to the investigator who was working with deaf or dumb subjects, or with subjects not acquainted with the language in which the tests were being made. The practical problem arising out of the need to measure the intelligence of non-English speaking immigrants into the United States was, as we saw, one of the reasons leading to the devising of the performance test.

The essential characteristic of the performance test is that it shall not require any kind of a language response on the part of the subject for an adequate performance of the test. Obviously it is unfair to expect to get an adequate response from a child who is not familiar with the language that is being used. In the United States there has been a great influx of population from non-English speaking countries, and it has been in the United States that the greatest amount of work has been done in the measurement of intelligence. So the problem of the foreign-born soon impressed itself on those who were working in the field of mental measurement. Other workers encountered difficulty with the Binet tests as they tried to use them with the deaf and with those defective in speech. Defective hearing and defective speech are physical defects in the first instance. There is no necessary connexion between mental and physical defects. It might be that an investigation would show that a larger percentage among the deaf and dumb are mentally deficient than among subjects of normal hearing and speech, but that would not alter the fact that the language test is inadequate. For many of those who are deaf and dumb are of quite good mental ability, but whether they are or not could never be discovered by a language test. It is only in exceptional cases that the deaf person sufficiently surmounts the language difficulty as to be able to respond well enough to be measured by such a standard.

Pintner and Paterson, who have done so much to develop the performance test, have been guided by three criteria in the selection of their tests. These criteria are related to three factors which must be taken into consideration, namely, first the complex character of intelligence, second the definition of intelligence adopted, and third the necessity of overcoming the language difficulty. It would seem that the first and second of these criteria are in reality two aspects of the same thing. In the second chapter we

have already dealt at some length with the problem of defining what it is that we are trying to measure by these tests. We have noted what Pintner and Paterson observe in their first criterion, viz., that intelligence is very complex. There are as many factors brought into play as enter into the constitution of a normal human being's conscious life. The variety of our responses to stimuli, the many-sidedness of our motives and intentions, and the breadth of our attitudes are illustrative of the complexity of conscious life. The complex character of intelligence means that it is very difficult to predict what a human being will do under specific circumstances. Of course the laws of habit make possible a certain amount of prediction, but a human being always has the possibility of inhibiting the habitual way of acting. It is the complexity of intelligence that enables a man to have the advantage over the lower animal in this matter of varying responses to stimuli. The lower animal is much more under the control of instinctive and habitual ways of responding than is the human being. The significance of all this for the psychological tests is that they must be so devised as to allow for the complexity of the mental processes, association, creative imagination, attention, or all the processes together.

The second criterion proposed by Pintner and Paterson is that the tests must measure the ability of the child to adapt himself to relatively new situations. This, as we observed before, is the definition accepted by these authors, as well as by Stern, of intelligence. Certainly it involves a factor of immense importance in the determination of a test. A test that involves only a familiar situation does not necessarily call for intelligence at all. If the response called for were familiar enough, it might be met automatically. If intelligence is to be tested, there must therefore be an element of novelty in the situation. To be sure, the terms novelty and familiarity are relatives and not absolutes. That would be equally true of a language test and of a performance test. The fact that a child may be familiar with certain words does not involve familiarity with the problem which they are utilized to express. So here the familiarity of the child with picture blocks does not militate strongly against them being used to express specific problems. On the other hand the devisers of performance tests have steadfastly avoided using anything for test material which is a plaything or toy with which children are very familiar. The process of perception itself includes elements both of familiarity and novelty. There must be a sufficient amount of familiarity to enable the person to identify or classify the experience or else it will not stimulate him to any perceptual experience. On the other hand there must be a change of some sort, some degree of novelty being presented, or else the person will from sheer fatigue cease to attend to the object of experience. The psychological

test which preserves just enough of the familiar to enable the subject to carry on a process of apperception, and at the same time presents a maximum of novelty, will at once command the interest of the subject, and, if it be a problem, will draw into play the creative processes of intelligence.

A third criterion which Pintner and Paterson set before themselves was that the tests should be so devised that they could be given and that the subjects could respond without the use of language. The obvious advantage of such a test is that it can be employed with subjects who use a foreign language and with those who are deaf or suffering from defective speech. Of course it would convey an impression of abnormality in the situation if an examiner said nothing, but gave his signals to proceed only in the form of gestures. On that account it is usual to give certain instructions in the case of children who can hear. But a performance test is so devised that it can be given just as well without verbal instructions, so that it serves its purpose with no verbal instructions, nor are the subjects put to any disadvantage who are simply signalled by a gesture to proceed.

The psychologists who worked in the American army made the first extensive use of group tests. Furthermore they devised a group test of the performance type. It was their desire to work out a scale that would be suitable for all the men who came for examination from all parts of the country. But some knew the English language well, while others knew it very inadequately and a few not at all. It was no easy task to arrange a scale that would measure by an equitable standard the intelligence of both illiterates and literates. Reference will be made later to these tests as 'group tests.' There were two scales arranged, called the 'Alpha' and the 'Beta' examinations. The former was for the literates; the latter was for the illiterates. But the Beta examination was "in effect, although not in strictness test for test, Alpha translated into pictorial form so that pantomime and demonstration may be substituted for written and oral directions."[1] The Beta scale was not exactly a scale of performance tests in the sense, for example, that Pintner and Paterson's scale is, but it is somewhat of a paper adaptation of a performance scale. It occupies a midway place, so to say, between the strictly performance test and the language test, and the fact that it can be given to subjects quite illiterate in the English language may justify reference to it in this connexion. In addition to these group tests, the Army psychologists also employed individual tests for doubtful cases. One of the scales used for individual testing was also a scale of performance tests which were devised to meet the exigencies of the military situations with which the men were confronted.

---

1 Army Mental Tests, pp. 16, 17.

I propose now to describe some of the actual performance tests, commenting on their usefulness and validity as we proceed.

I. THE FORM-BOARD.

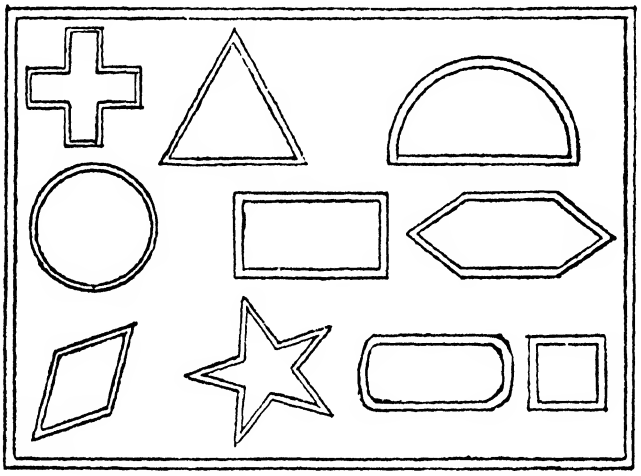
The best description of the essential features of a form-board is probably that given by Sylvester of the Seguin Form-board.[1] It runs as follows:—

"The ten geometrical figures, as nearly uniform in size as their variety of form will allow, are cut through an oak board 20 × 14 × ⅜ inches. This oak board is glued to a soft wood board of the same length and breadth, ⅝ inch thick. The result is a thick board of moderate weight with a hard oak surface in which the ten forms appear as shallow holes or recesses. About the edge is placed an oak strip, 1¼ × ¼ inches, fitting a ¼ inch raised edge about the oak surface. Corresponding to the ten recesses are ten walnut blocks, ⅞ inch in thickness, each of which fits loosely into its corresponding recess. The thickness being more than twice the depth of the recesses, the blocks can be easily grasped and removed. The board and the blocks are finished in their natural oak and walnut colours and the recesses are painted black. The whole is carefully finished in order to give it an attractive appearance—an important feature in a mental testing device. This description applies to what may be called the standard form-board—the type now in most general use."

The foregoing description, as I have indicated, is of the Seguin form-board. But, as it gives an indication of the general type and with the exception of a few details, it will suffice for any of the form-boards. The Goddard form-board is very much the same, so much so that Pintner and Paterson advise that the norms of either may be used for the other, the differences being only slight. To be sure Seguin first devised the form-board as an instrument for the training of feeble-minded children and its use as a device for mental measurement is a more recent development. The name of the device is significant, for it suggests that it calls for the perception of form to be successfully performed. The task is the perception of the different forms, either by sight or by touch, and making a definite movement of reaction with each form, namely placing it in its appropriate hole. Obviously sight and touch are the two factors that will be called into play the most. The test might be performed by means of either channel separately, but the two co-operating insure the best results. On the other hand, when we are dealing with older children and with adults, unless they are performing blindfolded in which case the perception of form operates very strongly, successful performance depends more largely on speed and co-ordination of movement. The use of the

[1] Psychological Monographs, Vol. XV, No. 4, Whole No. 65, 1913.

device as a measurement of intelligence involves an endeavour t make it a test rather of the perception of form than of speed an co-ordination of movement. That means that the administratio of the test must vary with the subjects under examinatior Happily the test lends itself to two methods the one for visual, an the other for tactual perception. The former is used largely fo the feeble-minded and for children of seven years or younge while the latter is used for older children and for adults. In cas of doubt the tactual is tried first.

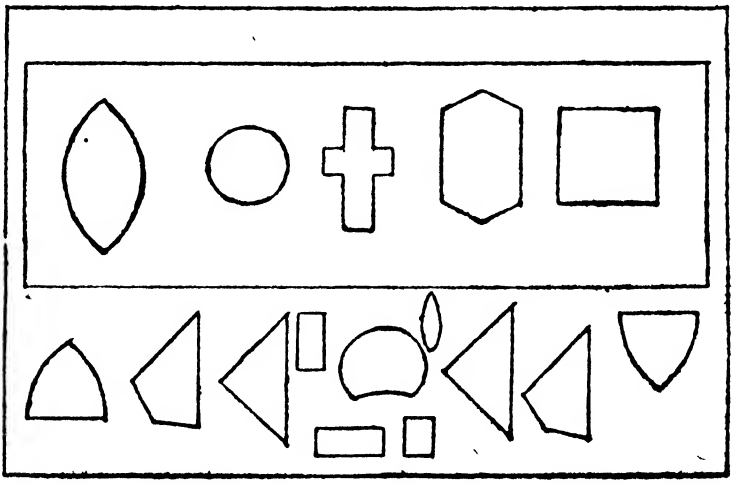


The Goddard Form-Board.

As to the method of procedure, we may quote again from Sylvester. He says: "The form-board lies horizontally on a table, its lower edge even with the edge of the table next to which the child stands. The table must be low enough to allow him to lean well over the board and to look down upon its centre. The blocks are placed in three piles on the table next to the upper edges of the board, no block in the pile nearest its recess, the lozenge and the elongated hexagon not in the same layer, and the star in the lower layer. This is the arrangement at the beginning of each of three trials. The child is introduced to the test with no introduction concerning it except, 'Let us see how quickly you can put the blocks in place.' His first reactions and his behaviour until he succeeds in getting the blocks into place or fails are carefully studied. After this first trial he is given any instruction necessary to make him understand where the blocks belong and that he is to replace them as quickly as possible. Then he is given a second and a third trial, in which he is encouraged and urged in every way to make the best record of which he is capable. These last two trials are timed with a stop watch and the shortest of the two records is taken as the child's form-board index."

In order to standardize the test still further, an arrangement of the blocks has been agreed upon by examiners, so that all subjects will start with the blocks in the same place. For left-handed subjects the arrangement is the reverse of that for right-handed subjects, and the same arrangement is used for the tactual method as that employed in the visual method. Variations in the method are suggested in Whipple's Manual. For example, the board may, without warning to the subject, be suddenly turned to a different angle when the subject is in the middle of the performance. Or, he may be allowed to make a visual study of the holes and blocks, and then be blindfolded, at the same time turning the board through 90 degrees. Or, he may be allowed to try first with one hand, then with the other, and finally with both. Another useful variant for use with adults is to have, as suggested by Mr. D. G. Fraser, a board in which the holes are made in a series of removable blocks of such dimensions as to permit of a free interchange within the board as a whole, thus allowing arrangements in different groupings.

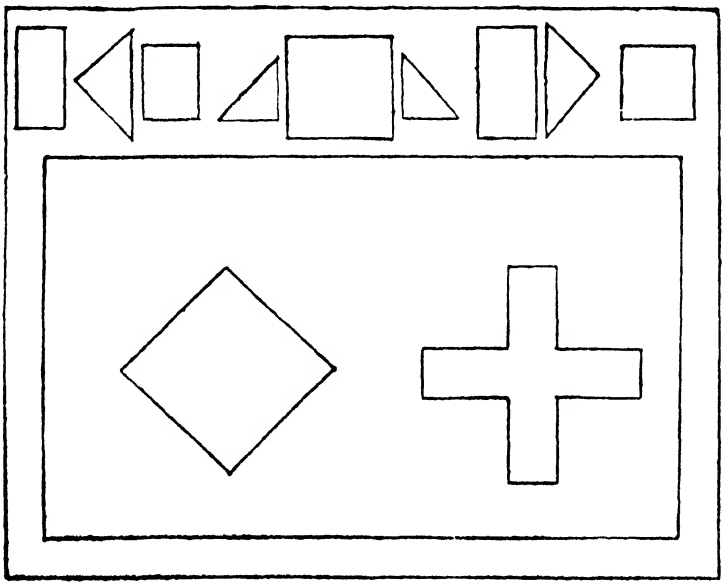
Two other types of form-boards were devised by Pintner and Paterson, and find a place in their scale of tests. One is a Two-Figure board and the other is a Five-Figure board, the former having been devised by Pintner and the latter by Paterson. In



The Five-Figure Form-Board.

these cases the cut-out blocks are divided into pieces so that the reconstruction is made more difficult. The Five-Figure board was devised to test a little higher grade of intelligence than the Seguin or Goddard form-boards. In it the blocks with one exception are divided into two pieces and that one into three, and the results confirm the experimenters in their opinion that it tests a slightly

higher grade of mentality. The Two-Figure board was intended to test still higher mentality. So two figures were used, the square and the cross, in the former case five pieces having to be put together to fill the recess, and in the latter case four. But the results show that it is somewhat easier than the Five-Figure form-board. In each case the method of procedure is much the same as has been described for use with the Goddard form-board. One

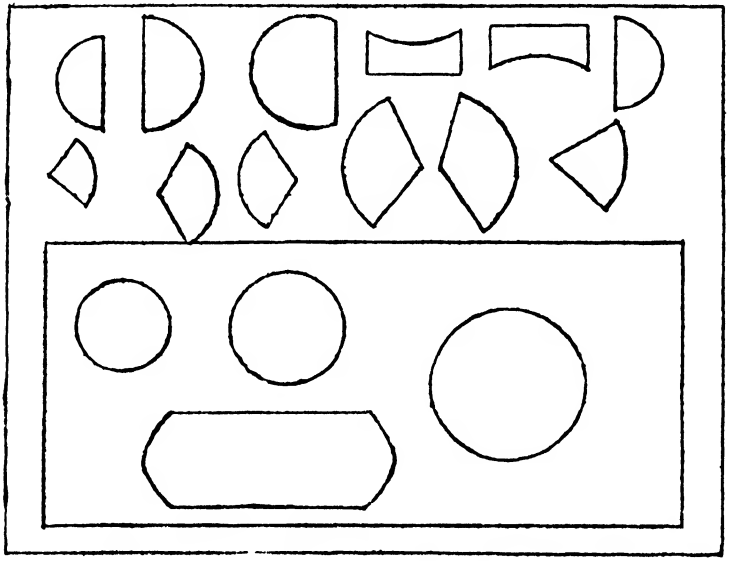


The Two-Figure Form-Board.

element is introduced into the scoring however which was not found necessary in the previous case—a record is kept of the number of errors made. An attempt to fit a block into a wrong hole constitutes an error, but not the holding of a piece above a wrong hole, if the subject does not try to insert it. In each case, as also with the Goddard form-board, a time limit of five minutes is fixed.

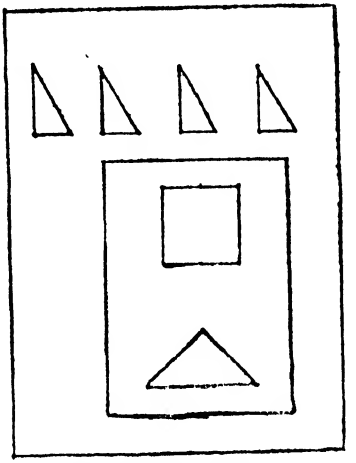
Another creation of the form-board type of test is the Casuist form-board which was devised by H. A. Knox in his work with immigrants at Ellis Island. The recesses in this instance are three circles of different sizes and an elongated oval with sides parallel for part of the way. The blocks for the two larger circles are cut into three segments each, that for the smaller circle into two equal segments, and that for the oval into four pieces. Knox standardized the test as one of twelve-year mentality with an allowance for what he calls "sensible mistakes". Pintner and Paterson think it too easy at that age and find that seventy-five per cent of seven-year-old children are able to succeed, although they make an average of thirty mistakes "which would probably not fulfil Knox's

requirement of 'sensible mistakes'." The method of procedure is of the same type as that used in the other form-board tests.



The Casuist Form-Board.

The Triangle test is another of the form-board type which Gwyn devised and Knox used in testing immigrants. There are two recesses in the board, one triangular and the other rectangular in shape. The rectangle is cut into two parts by a

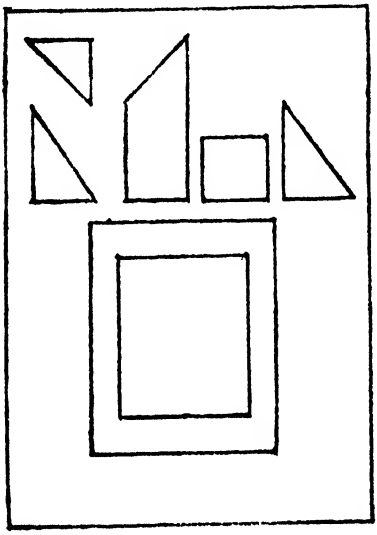


The Triangle Test.

diagonal cut, while the triangle is cut into two equal parts by a line running perpendicularly from the apex to the middle of the base line. This results in four triangular pieces which are exactly the same size. The method of procedure as before is to place the board and the blocks before the subject, asking him to put them

together as quickly as possible, five minutes being allowed for the trials.

The Diagonal test was devised by Kempf and also adopted by Knox. This test introduces a new element, inasmuch as there are two or three possible solutions. The chance factor thus enters into the attempts at solutions, and moreover the different solutions attainable are not equally difficult, so that if a subject happens to begin with one of the easier methods he has a better chance than the subject who begins with one of the more difficult solutions. We might describe the test as a combination of form-board and puzzle. It is not so significant as to how the blocks are arranged in this tests, so long as adjacent pieces are not placed contiguously, since there is more than one way of doing the performance. An error is recorded if the subject introduces a block in such a way that the other blocks could not possibly be fitted in, a fact which considerably reduces the number of errors, because of the various possible arrangements.



The Diagonal Test.

Another type of the form-board test was employed by the American Army psychologists. In this experiment blocks of various shapes are used—squares, triangles, circles, half-circles, and so on. An arrangement of the blocks is made by the examiner which leaves out, in the first problem, a square which cannot be fitted into the remaining recesses. The subject is then asked to rearrange the blocks in the fewest possible moves so that the square can be put in place and no blocks will be left over. Before setting the problem, however, a demonstration problem is shown to the subject by the examiner. In a second problem the subject is required to find places for two extra squares, and in a third problem places have to be found for four extra blocks. The time limit for the first two problems is two minutes each, and for the third three minutes. A scale of marking was standardized on the basis of the number of moves which the subject required in reaching a correct solution, a move being defined as "placing or trying to place a block in some position on the board." In the case of non-English speaking subjects the examiners gave their instructions by gestures only.

It may be pointed out that the form-board examination is a test of two factors, the one quantitative and the other qualitative. The quantitative element is indicated by the speed of the performance, the first trial being taken as the measure of the subject's normal unpracticed performance, where there is no disturbing factor for which allowance must be made. The qualitative element is indicated by the number of errors in the performance, an error being regarded as an index to the subject's inability to perceive or to recognize form. Where the visual method is employed, Whipple remarks that "persistent attempts to insert a block where it is manifestly impossible for it to go, or such absurd things as turning the blocks upside down to make them fit, standing them on end, etc., should be especially noted, as they are symptomatic of decided immaturity and are often seen in mentally defective subjects".[1]

One of the most interesting experiments with the form-board, as far as we are concerned, was that under taken by the Rev. D. S. Herrick, M.A. of Bangalore.[2] Mr. Herrick examined over 700 children of all ages from four to fourteen and tabulated the results. Three hundred and fifty-five children were Panchamas and 355 were Brahmans in 20 or more schools in this Presidency. Mr. Herrick says: "Not one of the more than 700 boys and girls tested had ever seen a form-board, it is safe to assert. Few, if any, of them in all probability had ever handled blocks of wood or other material of different shapes, much less tried to fit them into holes of corresponding shapes. To be confronted with the board full of holes and a lot of blocks, and to be told to put the blocks into the holes as quickly as possible, was a new situation for each of those children. Thus it was well adapted to test their intelligence. At the same time there was nothing unreasonable in the test, so perfectly simple is it."

In each case where the test was given there was an Indian teacher present to make sure that the language of the examiner was comprehended. In cases of doubt he repeated the command. In each instance three trials were given, both time and the errors being recorded. "The time of the fastest performance was regarded as the index of the subject's psycho-motor ability. In practice it was found best not to ask for speed at the first trial, as that tended to confuse him, and sometimes resulted in wild dashes at the board with little effort to avoid errors. A correct performance was the first thing aimed at. Before the second trial, however, the subject was told to put the blocks in as quickly as possible. Before the third he was urged to his utmost effect for greater speed." Mr. Herrick was careful to do his utmost to standardize the condition

---

[1] Manual of Mental and Physical Tests, Vol. I, p. 302.

[2] A comparison of Brahman and Panchama children in South India with each other and with American children by means of the Goddard Form-Board, printed in the *Journal of Applied Psychology*, September 1921.

under which the tests were given so that his comparisons might be made in fairness to all the subjects concerned. Performances which took more than five minutes were not recorded as such delay points to a defective mentality which it is unfair to include in comparing two groups.

The results of the experiment are of interest. On the average the Panchama child took two and one-half seconds longer than the Brahman child for the performance, a difference certainly not great in a test for which five minutes is allowed. Mr. Herrick thinks that this difference can perhaps be accounted for "by the great difference in social and educational opportunities enjoyed by the two groups in the past, and by the difference in their environment". Going on with the comparisons, Mr. Herrick observes that the Brahman child at four years is much quicker than the American child, the median times for the two groups being 41 as against 46 seconds. At five years, however, the American children catch up, and the median for both groups is 37 seconds. At six years the American children have improved to a median of 26 seconds while the Brahmans stand at 33 seconds. From that point onwards the American average continues to be from five to eight seconds better than the Brahman. Mr. Herrick in seeking for an explanation of this deviation, alludes to the fact that climatic conditions greatly affect the rate of maturing among children. He wisely suggests that when Indian education makes larger use of the kindergarten with its training in free manipulation there should be an improvement in ability to respond to tests of this type. At the same time, we must all admit that the numbers so far tested have not been sufficient for any broad generalizations, though the results so far obtained are full of interest and suggestion.

2. THE PICTURE FORM-BOARD.

A number of tests of the general character of picture form-boards have been devised. These vary from the tests which have been described in that they make use of pictures that have to be reconstructed instead of geometrical figures. The subject is required to insert blocks in recesses to which they correspond. Substantially the same mental processes are brought into play as in the case of the other form-boards which are made with geometrical figures. The following noteworthy passage in Whipple describes the mental processes: "This complexity in the mental processes concerned in the tests is reflected in the statements of those who have made most use of it. Norsworthy, for instance, called it a 'test of form perception and rate of movement,' and also sought to secure indication of learning capacity from her data. Jones likewise used the test to determine learning capacity, and speaks of it, too, as 'a very good test of native ability'. This idea that the test has diagnostic value in examining intelligence is again reflected in

Norsworthy's statement that 'this test seems to me to measure to a certain extent the ability of dealing quickly and well with a new situation' (which approximates Stern's definition of intelligence), and in Witmer's statement that 'the form-board is one of the best tests rapidly to distinguish between the feeble-minded and the normal child,' to which he adds that 'it very quickly gives the experimenter a general idea of the child's powers of recognition, discrimination, memory, and co-ordination', while 'repetition of the experiment leads to a conclusion as to his ability to learn'. Wallin believes that the form-board test throws light upon the patient's ability to identify forms visually, upon his constructive capacity and his power of muscular co-ordination. Goddard says: 'We have in our laboratory no other test that shows us so much about a child's condition in so short a time.' His table of norms suggests strongly that the test can be of direct service in the examination and classification of mentally defective children."[1]

The Mare and Foal Picture Board is one of the picture form-boards. It was originally devised by Healy after which a modification was made by Pintner and Paterson. It consists of a board about $9\frac{1}{2} \times 11\frac{1}{2}$ inches upon which a coloured picture is pasted. The picture is of a mare and her foal in a field with two sheep lying down and three chickens in the foreground. Two houses are to be seen in the distant background. From the whole picture eleven pieces have been cut, differing in shape and size, and representing parts of animals or of the scene. The original Healy form of the picture had four geometrical forms inserted in the top part of the picture which have been omitted from the Pintner and Paterson modification, first because they differ so radically from the test as a whole and in the second place because the other form-board tests, particularly the triangle test, call for all that is demanded by this additional feature in the Mare and Foal Test. The modified test seems much less likely to confuse the child, and it would appear to be wiser to test the different abilities separately. In the case of the pictures the child will find guidance in the cut-out as well as in the shape. The method of procedure resembles that of the other form-board tests. The child has the frame and the pieces placed before him, and is asked to put the pictures into their appropriate places as rapidly as possible without making any errors. The performer is timed by a stop-watch and at the same time his errors are recorded by the examiner. An error is any attempt to make a piece fit into a wrong space, but the holding of a piece above a space is not recorded as an error, if the child does not deliberately try to make it fit. Five minutes is the time limit of the test. This test, with certain modifications to make the scene typically Indian, ought to prove to be a valuable test for use in this country.

---

[1] Op. cit., Vol. I, p. 297.

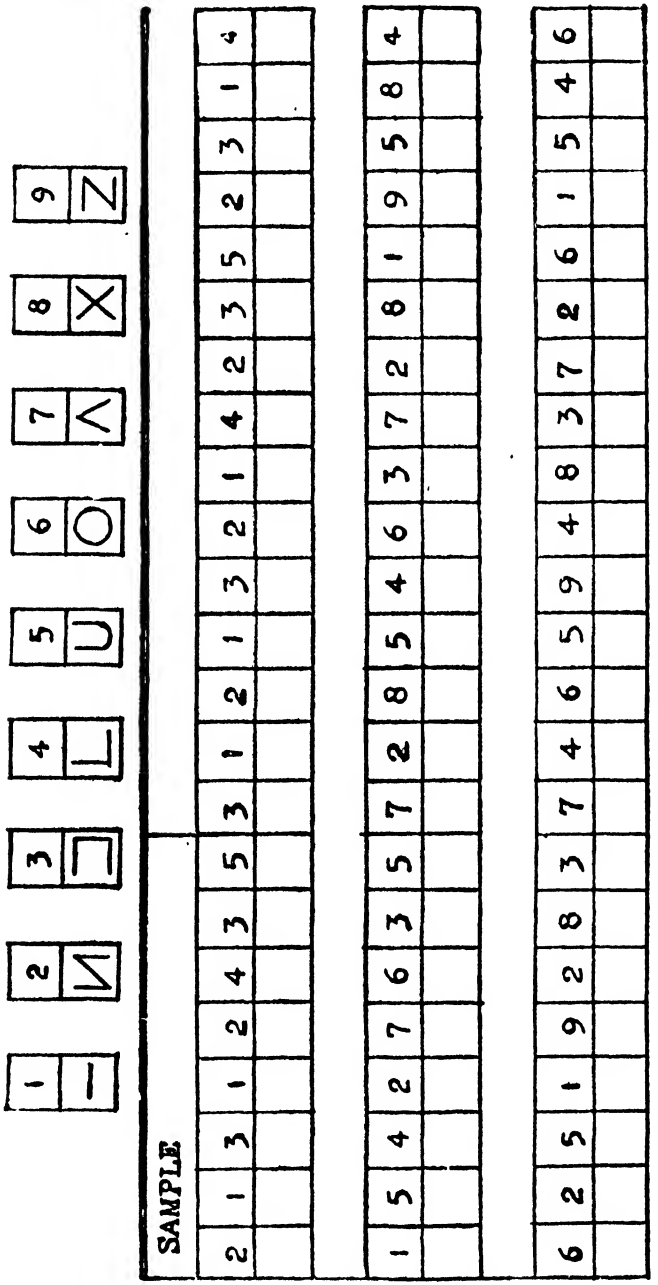
The Ship Test follows the same general plan as the Mare and Foal Test, but it has this difference that all the pieces are the same size and shape. Gluck has the merit of having devised the test, Knox used it, so did Pintner and Paterson, and lastly it is included in the performance scale of the Army Mental Tests. The size and shape of the pieces will be no assistance in determining the places they must occupy. The subject must be guided solely by the picture which he is making, an objective which varies in coherence. It will be quite apparent that there will be all varieties of performance from one that is perfectly coherent to one that is absolutely meaningless. Hence the scoring has to be so arranged as to take into account the different grades of correctness. The methods of scoring as suggested by Pintner and Paterson and that used by the Army psychologists were different, but had this in common that they made provision for a graded scoring in accordance with the measure of correctness which the subject attained. The Army men put a time limit of five minutes and gave marks for speed as well as for accuracy. Pintner and Paterson suggest no time limit, though they note that 60 per cent of thirteen-year-olds complete the test within five minutes. During the performance the subject is allowed to make as many corrections as he choses without losing credit. Indeed the test is especially useful in testing those abilities of devising means to an end as well as of auto-criticism which Binet noted as characteristic of the function of intelligence.

Another test, which is a development of the form-board, making it still more complicated, is the Picture Completion Test. The test is in the form of a picture or a series of pictures from which certain features are missing. In addition a large selection of smaller pictures are provided of the same size as the empty places in the larger picture which empty places, it may be observed, are of uniform size. The subject has the larger picture placed before him, as well as the smaller ones in heterogeneous order, and he is asked to select from the smaller ones the appropriate ones to complete the larger ones. Pintner and Paterson have a test of this type which they have adopted from Pintner and Anderson. Healy has also a Picture Completion Test. The Army Mental Tests also included a test of the same type, the difference being that the former is a single picture and the latter a series. The time limit by Pintner and Anderson and by the Army psychologists suggested is ten minutes, whereas Healy placed a limit of five minutes. It will be apparent that this is a special application of the completion test fathered by Ebbinghaus, and we have already commented[1] upon the method involved as calling forth fundamental processes of intelligence and correlating highly with other tests of intelligence.

Another type of performance test is the substitution test. "This test," as Whipple says, "is one of many that may be devised to measure the rapidity with which new associations are formed by repetitions. The name commonly applied to the test arises from the process that it involves, in which the subject is called upon to substitute for one set of characters (letters, digits familiar geometrical forms, etc.) another set of characters in accordance with a plan set before him in a printed key. The procedure differs from most memory tests or exercises of memorizing in that the connections indicated by the key are not committed to memory at the outset, but acquired gradually by use as the test proceeds." A number of variations of the substitution test have been employed by different investigators, especially in connexion with the study of the psychology or learning.

An example of the substitution test which has been widely used is the Digit-Symbol Test. Whipple, Woolley and Fischer Woodworth and Wells, Baldwin, Pyle and others have all made use of it in some form. The Woodworth-Wells form was adopted by Pintner and Paterson. The Army psychologists followed the lead of Whipple. The Whipple test is to place before the subject a card on which there are nine circles in each of which there is a number from 1 to 9, and a small figure or drawing. Then he is given a strip of paper with rows of the same character and with empty squares beside them. The subject is then told that he is expected to write in the empty squares the numbers corresponding to the figures and to continue persistently until all the empty squares have been so filled in. The army test reverses the process. That is to say, the strips contain the numbers and the subject is to fill in the corresponding characters. The Woodworth-Wells test contains five figures of different shapes—star, circle, squares maltese cross, and triangle, each of which has a number. The strips of paper contain rows of these figures and the subject is asked to insert the appropriate number in the figures throughout the strips. The examiner observes the number of errors made, the time taken for the entire test, the gain made towards the last as related to the speed of the subject at the beginning of the performance, the accuracy of the performance, and the knowledge of the symbols. Woodworth and Wells suggested that the penalty for each error be fixed in ratio to the total time occupied to complete the test, each error being scored as 1/50th of the total time for the test. The method was reached on the theory that, were the child afforded an opportunity to correct his mistakes, the actual time for correcting them would be equivalent to the time occupied in filling in one figure. Investigators have found that the substitution test correlates positively and highly with intelligence. In the case of delinquents who were tested they were able to perform correctly but required a much longer time than normals, whereas in

the case of mental defectives the success attained was much poorer and the time occupied much greater. Though schooling



Digit-Symbol Test used by American Army Psychologists.

undoubtedly helps in the attainment of success, still it does not function so largely as does intelligence,

Cubes are used in a variety of performance tests by various investigators. Knox devised one which has been adopted and standardized by Pintner. In this test five blocks of the same size and shape are used, four of which are placed in a row in front of the subject at a distance of about two inches from one another. The examiner takes the fifth cube and taps on the other four in different combinations, and the subject is asked to do exactly as the examiner has done, the examiner recording the number of lines done correctly and number done incorrectly. The test shows a satisfactory distribution, the very young sometimes failing completely, and the number of correct performances increasing with advancing age. The American Army psychologists made use of quite a different test involving construction instead of tapping according to a definite arrangement. Problems of construction were assigned to the subject, and he was judged according to speed, the number of moves he made, and correctness of assemblage. The test in whichever form it be used is obviously one which calls into function the associative processes as well as the power of auto-criticism.

Another well-known example of the performance test is the Maze Test. The maze is of interest because of its use in animal psychology to measure the animal's ability to learn. In human psychology it has been made to serve various purposes, as tests of learning ability, of attention, and of perception. Whipple describes an attempt made by Burnett to use the maze to measure visual attention. He employed two mazes that were alike except that small pictures and bits of paper were scattered among the twistings of the maze, although not actually concealing any portion of it. The measure of attention is taken by the time taken in maze one where there is no distraction as compared with that taken in maze two where there is distraction, in a limited number of trials. Burnett ascertained that the distraction was not too great to be overcome by adult intelligence. In fact the extra effort so called forth results in an increase rather than a decrease in the speed of tracing.

The American Army psychologists made use of the Maze Test with four problems of that type. It was also employed in the Group Test Beta, as indeed it is in other group tests. We shall find not only the Maze Test, but other performance tests recurring in the group tests.

The comparison of the performance of animals with humans in the Maze Test is illuminating. Woodworth gives the following table, showing the number of errors made in successive performances of white rats, children and adults. This of course is for the actual threading of a maze and not simply for the tracing of one on paper, and therefore involves more of the learning process with less opportunity for relying on visual perception.

Trial Number.	Rate.	Children.	Adult Men.
1	53	35	10
2	45	9	15
3	30	18	5
4	22	11	2
5	11	9	6
6	8	13	4
7	9	6	2
8	4	6	2
9	9	5	1
10	3	5	1
11	4	1	0
12	5	0	1
13	4	1	1
14	4	0	1
15	4	1	1
16	2	0	1
17	1	0	1

(Table from Hicks and Carr.)[1]

The method of scoring and of arriving at a measure of intelligence in accordance with a scale is a problem which confronts those who use a performance scale. The Army psychologists were governed by a particularized motive. Their criterion was military efficiency, and the intelligence measurements were means to such an end. They did not confine themselves to any one scale of tests, but employed group tests of both the language and performance types as well as individual tests of both types. It was necessary to have a method of scoring which would yield standardized results in dealing with such large numbers of subjects by different methods. They expressed their different classes of intelligence by means of letter grades and had a system of credits for the various tests, including the performance tests, which they converted into the letter grades. The following table, taken from *Army Mental Tests*, (p. 17), indicates the method employed:—

Intelligence Grade.	Definition.	Score (Alpha)	Score (Beta).
A	Very superior	135-212	100-118
B	Superior	105-134	90-99
C +	High average	75-104	80-89
C	Average	45-74	65-79
C -	Low average	25-44	45-64
D	Inferior	15-24	20-44
D	Very inferior	0-14	0-19

Pintner and Paterson have summarized the results of their investigations in a very useful way. I cannot do better than quote from their summary, in conclusion.

---

[1] Woodworth, R S.: Psychology: A Study of Mental Life, p. 314.

"1. A scale of performance tests as a means of estimating mentality is needed for those children who are deficient or wanting in language.

2. Such a scale is the only means that can be used to measure the intelligence of the deaf, the speech defective and the non-English speaking individual.

3. Language ability is not uniformly correlated with general intelligence, and therefore a scale of performance tests will be a useful supplement to other scales which depend entirely or in part upon language responses.

4. The need for a more adequate standardization of most of the performance tests in common use has led to an effort on our part to supply this deficiency.

5. The value of such performance tests is greatly enhanced when they are grouped together in some kind of a scale.

6. The results of the tests are presented in tables of distribution so that additional results may be added from time to time and the reliability of the norms thereby increased.

7. Four different methods of arriving at an index of mental ability have been discussed.

8. The year scale method has the advantage of leading to a result that is easy to interpret, but it has the disadvantage of requiring a great many different tests. This would make the scale unwieldy and would lengthen, beyond practical limits, the time taken to examine a case.

9. We have attempted to construct with our tests a modified type of year scale. This type of year scale differs somewhat from the type of year scale in common use. This difference is necessary if we are to overcome the disadvantages in the year scale method mentioned in the preceding section.

10. The median mental age method is simple in computation and permits the addition or subtraction of tests without dislocating the whole scale. Difficulties arise when the medians are the same for several consecutive ages. The diagnostic significance of the median mental age is yet to be determined.

11. The point scale method has been subjected to a discussion in order to find out the most satisfactory underlying principle upon which to base a point scale. The results seem to lead back to a method clearly akin to the median mental age method and showing no superiority over that method.

12. A point scale has been constructed on the principle of the allotment of the same number of points to each test, although the value of this method of procedure is doubtful.

13. The percentile method seems to offer the best possibilities for future work. The percentile division can be made as small as

the delicacy of the tests will warrant. This method is especially desirable because it permits us to compare an individual's performance with the performances of other individuals of the same age. It would seem at present, however, to require for purposes of standardization, a very great number of unselected individuals at each age.

14. These different methods lead to different estimates of mentality for the same individual. Which leads to the truest estimate of intelligence is a problem still to be solved.

15. The correlation of this scale with scales of the Yerkes or Binet type has not yet been attempted. Whether a scale of performance tests or a mixed scale of performance and language tests will yield the best estimate of intelligence has yet to be determined." [1]

---

[1] Op. cit., chapter X, pp. 210 ff.

CHAPTER VI.

GROUP TESTS OF INTELLIGENCE.

The exigencies of military training were responsible for the first extensive use of group tests of intelligence. There had been some scattered experiments in the direction of group tests, but there was nothing uniform or systematic. But when the psychologists of the United States mobilized for war service they at once appreciated the need for group tests. Men were brought into the army rapidly and came in large numbers to the training camps. The mental rating of a man to be of the best service to the army should be available as early as possible after the man enters the training camp. If there were no other method available than that of the individual tests, which take from forty minutes to an hour to administer to each individual, it would require a small army of psychologists to test the larger army of enlisted men. So in the interests of the economy of time the group test for the measurement of intelligence simply had to be developed, if intelligence tests were to be of any practical value to the army.

The committee of psychologists which first met to consider what services could be rendered to the army outlined the following conditions for tests that might be made available for army use in the examination of its personnel:—

(i) A test should be adaptable for group use in the examination of large numbers of men rapidly.

(ii) It should possess a high degree of validity as a measure of intelligence.

(iii) The tests should be capable of measuring a wide range of intelligence, including the highest and the lower levels.

(iv) The scale should be arranged for objectivity of scoring and the elimination of personal opinion, thus preserving the advantages of standardization.

(v) The tests should be arranged so that the examiner can score the results with a maximum of rapidity and a minimum of error. Moreover the arrangement for scoring should be such that examiners might make use of relatively inexpert assistants. This corresponds to what Ballard emphasizes as a necessary factor in insisting that the tests must be "fool-proof."

(vi) To avoid coaching, a variety of forms or alternative must be available.

(vii) Clues are necessary to assist the examiners in detecting subjects who may sham illness to avoid taking the test.

(viii) There must be a minimum of opportunity for cheating.

(ix) Tests must be made as far as possible independent of schooling and educational advantages.

(x) The arrangement should be such as to call for a minimun of written responses.

(xi) The tests should be designed with reference to arousing the interest of the subjects.

(xii) The arrangement of the tests should be such as to enable the examiners to secure an accurate measure of the intelligence of the subjects in the shortest possible time.

The above were the criteria which the examiners had in mind in the selection of tests for army use. There were a number of tests available when they began their work. Some of these were in printed form; others were in manuscript. A careful selection was made from the available tests and these were arranged, as we have already noted, in two groups, the former called Alpha and the latter Beta, the Beta being as nearly as possible a performance counterpart of the Alpha scale. These scales were then put to the test by applying them to approximately 80,000 men in four army cantonments and to about 7,000 college, high school and elementary school students. The data made available by these trial tests was then subjected to statistical treatment for the revision and standardization of the tests. It was a great array of experts who co-operated in this work, and for two months they studied together the results, checking them with all manner of available data.

"An Examiner's Guide for Psychological Examining in the Army" was prepared in which were contained directions for examiners who gave the tests. An introductory statement summarized the purposes of psychological examination with special reference to the military situation. The general plan of the examination with instructions for organization and routine came next. Emphasis was placed on the following points:—

(i) An adequate system of arrangements whereby men should report to the psychological officers for examination as promptly as possible after admission to a camp was demanded.

(ii) Group and individual examination blanks had to be examined and the results reported with all possible promptness to the military officers. A complete file of records was maintained by the Psychological department.

(iii) The intelligence rating and comment on any special aptitude of each man was reported promptly to the personnel officer, whereas company commanders were also provided with all relevant information.

(iv) All instances of mental deficiency as well as cases needing neuro-psychiatric examination were reported at once to the camp surgeon for the information of the psychiatrist.

(v) The psychological record card with any recommendations regarding the disposition of the case were forwarded to the office of the Surgeon-General.

It was especially urged that the results of examination should be made available as early as possible to personnel officers and to line officers. The instructions read: "It is, therefore, the duty of the psychological examiner to see that every drafted man is examined as promptly as possible after arrival in camp, and that report is immediately made to the personnel officer, to the medical officer if the case requires it, and subsequently to the company commander to whom the man is assigned."[1]

It was repeatly urged that "to be of the greatest value the psychological examination should be given at the earliest possible date after the arrival of the men in camp, in order that the personnel officer may have the results on the qualification cards when making assignments. Unless the scores are available and used properly at the time, companies will be built up that are very uneven in general intelligence. In order to balance companies and regiments satisfactorily it is necesssary to observe not only the special requirements laid down in the tables of organization, but also the requirement that there shall be equivalent grades of intelligence in company organizations and in the various trades and occupations demanded in each."[2]

Obviously the attainment of the end which the Army psychologists set before themselves could never be realized by the use of individual tests on account of the time involved in their administration. Where hundreds of men were being examined and the reports were required in the shortest possible time, a group method of testing with a fairly mechanical means of recording scores was necessary. In the same period of time which it would take to administer a Binet test to one individual it is possible to give a group test to one or two hundred men. The Army psychologists allowed from fifteen minutes to an hour for the administration of a Stanford-Binet, a point-scale or a performance examination, 50 to 60 minutes for the administration of examination Beta, and 40 to 50 minutes for examination Alpha.

As already pointed out, when the committee first met to discuss what could be done, there were available many tests, some in print and some in manuscript from which they could select. Among them was a scale of Group tests devised by Professor A. S. Otis of Leland Stanford University. The Alpha scale which the Army adopted was modelled on the same principle as the Otis Group test. The Beta test was parallel to the Alpha, performance being substituted for language for the sake of illiterate subjects.

The question which naturally suggests itself to inquiring minds is, What validity and reliability do the group tests possess? This query is especially pertinent in comparing the usefulness and

---

[1] Army Mental Tests, p. 47. [2] *Ibid.*, p. 48.

reliability of the group tests with the individual tests. We want to know whether or not they give us as accurate a measurement, and if there is any variation in the accuracy whether or not we can calculate the limits within which the variation will fall. [The whole question will be discussed in Chapter IX.]

Such was the problem which the Army psychologists faced when they first began their experiments with the group tests. Since the group tests were quite new in the field of educational psychology, there was no available data. They had to pave a way for their own advance. At the close of the eighteen months of service, however, they had accumulated a mass of material, on the basis of which, they have been able to measure the validity of the group method of testing, and to compare it with the individual method, until then in vogue.

During the period of investigation the group tests were given to 1,726,966 men of whom 41,000 were officers, and the individual tests were given to 83,000 men. This volume of data constitutes in itself as safe a basis, on which to calculate validity, as anything thus far available from other sources. Yoakum and Yerkes give us the following statistics which are valuable, not simply as a record of an interesting piece of work, but also as a guide to the general degree of validity which the group tests possess :—

" For examination Alpha the probable error of the score is approximately five points. This is one-eighth of the standard deviation of the score distribution for unselected soldiers. The reliability co-efficient is approximately .95. Alpha yields correlations with other measures of intelligence as follows: (1) with officers' ratings of their men, .50 to .70; (2) with Stanford-Binet measurements, .80 to .90; (3) with Trabue B and C completion tests combined, .72; (4) with examination Beta, .80; (5) with composite of Alpha, Beta and Stanford-Binet, .94; (6) in the case of school children Alpha measurements correlate with (*a*) teachers' ratings, .67 to .82, (*b*) school marks, .50 to .60, (*c*) school grade location, of thirteen and fourteen-year-old pupils, .75 to .91, (*d*) age of pupils .83.

" Results for examination Beta correlate with Alpha, .80; with Stanford-Binet, .73; with composite of Alpha, Beta and Stanford-Binet, .91.

" Results of repetition of the Stanford-Binet examination in the case of school children correlate .94 to .97. The abbreviated form of the Stanford-Binet scale consisting of only two tests per year, extensively used in the army, correlates .92 with results for the entire scale.

" Reliability co-efficients for results of point scale examination closely approximate those for the Stanford-Binet scale.

" The several tests of the performance scale, taken separately, correlate with Standford-Binet measurements, .48 to .78. Five of

the ten tests of the performance scale yield a total score which correlates ·84 with Stanford-Binet results.

"It is definitely established that examination Alpha measures literate men very satisfactorily, considering the time required, for mental ages above eleven years. Examination Beta is somewhat less accurate than Alpha for the higher ranges of intelligence. There are convincing evidences that some men are not fairly measured by either Alpha or Beta and that the provision of careful individual examination for men who fail in Beta is therefore of extreme importance."[1]

There has been a good number of scholars busily at work in this particular field since the work was given such an impetus by the Army psychologists. A number of group tests are now available of which the following may be mentioned as among the more important: the Otis Group Intelligence Scale, the Trabue Language Scales, Haggerty's Intelligence Examination Delta 1 and Delta 2, Whipple's Group Tests for the Grammar Grades, Myers' Mental Measure, Pressey Cross-Out Tests, Detroit First-grade Intelligence Test, Indiana Mental Survey, Dearborn Group Intelligence Tests Terman's Group Test of Mental Ability, Kingsbury Primary Group Intelligence Scale, The Simplex Group Intelligence Scale, The Miller Mental Ability Test, The Thorndike Intelligence Examination for High School Graduates, The Thurstone Psychological Examination for College Freshman and High School Seniors, The Northumberland Mental Tests, The Chelsea Mental Tests, The Columbian Mental Tests, Roback's Mentality Tests for Superior Adults and the National Intelligence Tests. The latter was prepared by a group of psychologists including Yerkes, Thorndike, Whipple, Terman, and Haggerty under the auspices of the National Research Council of the United States and is an application of the army testing methods to school needs. In addition to the tests mentioned there are others, some of which are devised with reference to some special needs. Here in India there have been some educationalists who have been devising tests with reference to the specific conditions which prevail in this country. The Narsinghpur Tests used in the Methodist Episcopal High School, Narsinghpur, Central Provinces and the General Intelligence Test used by the Cushing High School, Rangoon, are two adaptations to Indian conditions which have been used with a fair degree of satisfaction to the educators who have arranged them. Other experiments are being conducted in various parts of the country, including attempts at adaptation of the Terman Group Test, the Whipple Tests, and possibly others which have not come under my observation. It is too early to prognosticate as to which form of group test will be found the most adaptable to Indian conditions. Perhaps no one scale will

[1] Army Mental Tests, pp. 20, 21.

be found to meet the needs in all parts of the country. But the number of scales in existence seems to indicate that there is no unanimity yet in the countries where the investigations have been carried on the longest. When we consider that the Group Test as a measure of intelligence is scarcely more than five years old, it is probably too much to expect unanimity yet. We are really in the experimental stage. But it is important in the stage of experimentation that there should be as much data as possible brought to light from various parts of the world in the interests of standardization. If we are to be able to make comparisons regarding intelligence, either in a general way or in any of its constituent elements, or regarding the progress and achievements of subjects in different parts of the world, it is necessary that there should be some recognized standard by which we can make the calculations. Such a standard can only be formulated as workers in different areas experiment, and the cumulative results are subjected to careful scrutiny. It is important then that we in India be alive to the problems now, while we are still in the experimental stage.

When an intelligence survey of a school is to be made by the group method of testing, it is necessary to make careful preparations. Obviously, as in the case of the army, in a school the group method has the advantage that a whole class can be given a test simultaneously. This avoids the possibility of children who have taken a test coaching others. If the same scale be used in various schools, of course, it is possible to make comparisons of various classes of the same school grade. This is actually done in many cities where the educational authorities decide to test all of the children in all of their schools by a certain scale. Although the intelligence test does not afford a basis for organization comparable to the achievement test, still it affords the data for a comparative study of the intelligence of children in various parts of a community and frequently suggests the causes for certain disparities in other examination results. In the Madras Presidency we are familiar with the spectacle of one school persistently producing better results than another within the same municipal or union limits. It is quite possible that the subjection of the two schools to the same examination would throw some light on the difference, because the one school attracts the pupils of a higher grade of intelligence than the other. No amount of theorizing can answer the question. Investigation by actual experimenting alone can give the information desired.

"If a given school system is to have an intelligence survey," say the Myers, "detailed preparation should be made quietly after the fashion of getting ready to 'go over the top.' Let the Superintendent, or an expert designated by him, coach the principal and those of the teachers selected to give the tests. Let every teacher be imbued with the idea that the directions are to be

followed to the letter and that in order 'to put over' these directions each tester must be very familiar with them and with the process of precise reading of 'seconds' on a watch. Accurate timing of each test is of the greatest importance." [1]

It is important that the children who are to take the test be made as comfortable as possible, that they should be so put at their ease that they may do their normal best. They ought to be made to feel that they are co-operating with the examiners in a common task rather than they are being subjected to a sort of mental scrutiny. In the lower grades the tests may be presented in the form of puzzles or games; in the upper grades where the children have attained a measure of loyalty towards the school their enlistment may be secured in trying to make a record which will do credit to the institution. The greatest care must always be taken to prevent anything which will be in the nature of a disturbance, preventing the child from showing his real mental abilities. It is frequently thought advisable to have the examination conducted by instructors other than the regular teachers since the regular teachers, in spite of attempts to do otherwise, are liable to give little advantages to their own classes. Two or three seconds additional in the case of each test may seem a small matter, but it may amount to a half-minute on the whole test, and that amount would be ample to explain a few points of difference between the scores of two groups.

In the cases of group tests, the subjects are given printed tests with blanks left for their answers. In each case the front page contains blanks for the subject to fill in general information about himself which the examiner will want to have, such as name, whether a boy or girl, the grade in school, the subject's standing within the grade (to be secured from the school records), his age, last birthday, the date of his birthday, his nationality, the name of the school, the name of the teacher, the date of the examination, the name and occupation of the subject's father, his residence (which gives a clue to the social environment from which the subject comes), whether the subject is looking forward to any definite occupation, whether there are any points to be noted in regard to the subject's physical condition (such as deafness, defective eyesight, the presence of adenoids, etc.). Every blank does not call for all of this information, but I have selected points from various forms to show the kind of information in which examiners are interested. In addition the first page of the blank sometimes contains a score form in which the examiner records the subject's scores in the various tests and his total score. When the subject is given this blank he is instructed to fill in certain portions of it (the class works together), certain portions are afterwards obtained

[1] Measuring Minds, pp. 9, 10.

from the teacher's records, and certain parts such as the date and hour of the examination are filled in from the dictation of the examiner. Always the subject is instructed and carefully warned that the page must not be turned until the examiner gives the signal to do so. Before giving such instruction the examiner briefly and plainly explains the nature of the test, directs the class particularly to observe the printed instructions at the beginning of each test, and to do exactly what it is asked to do after the manner of the printed sample. He also instructs the class as to the time limit of the test and the necessity of stopping though they may be in the middle of a letter when the examiner calls time. The timing must be carefully done with a stop-watch, for standardization of such tests depends for one thing on an absotely common time element.

The American group tests with practically no exception insist on a time limit for the reason that time is one of the elements which needs to be standardized and that speed is one of the factors with which to reckon in intelligence. It has sometimes been objected that these tests lay too much stress on this factor, and that there is not sufficient time allowed for the one who may be a little bit slower worker and yet intelligent, to do himself justice. The British workers are more inclined to give the subjects a chance of working out their best without time limitations. Ballard expresses the criticism of the American plan in a cautious manner as follows:—

"The most serious criticism that has been made against the American group tests is that they put a premium on smartness—that they pick out the rapid thinkers and leave behind the profound thinkers. Those who devised the tests look upon brain-power just as engineers look upon horse-power: they regard it as a thing to be measured by the amount of work it can do in a given time. And this indeed is inevitable if we consider intelligence as including the ability to deal expeditiously with certain common tasks. Even Binet set time limits to some of his tests. For instance, in his counting test for eight-year-olds ('counting backwards from 20 to 1)' he allows only 20 seconds. He gives a child of twelve only one minute to rearrange the mixed sentence, 'a defends master dog good bravely his.' It is clear that if unlimited time were allowed, such questions would lose in distributive and diagnostic value. The valid objection is not that some of the army tests have time limits, but that all of them have time limits—that they contain no tests at all which give an equitable chance to the slow, cautious, and solid thinker. It is to meet this objection that in my own group tests some, if not all, of the questions are to be worked at the candidate's own pace[1]."

[1] Ballard: Group Tests of Intelligence, pp. 8, 9.

There is a large variety of tests which may be used and are being used in the group tests of intelligence. In going through the various lists to which I have had access I find the following tests in use:—the completion tests in various forms both of pictures and sentences, following directions, the opposites test, test of similarities, the rearrangement of dissected sentences, proverbs, arithmetical tests of various forms, geometric figures, the analogies test, tests of logical memory, tests of logical selection, correcting absurdities, copying designs, making comparisons, the symbol-digit test, cipher or code tests, orientation tests, tests of practical judgement, the synonym-antonym test, information tests, maze tests, classification tests, the true-false test, tests of meaning in sentences and paragraphs, the genus-species test, the part-whole test, and the number series test. We shall not have the time to examine all of these tests, and must select some of the more frequently used.

It will be apparent at once that some of the tests used in the group examinations are of the same types as those already considered in connexion with the individual examinations both through the language medium and through the medium of performance. For example, we have given some attention to the use of the completion method[1] which was first devised by Ebbinghaus, and have observed that it was used in several tests both of the language and the performance types. I find that there is no general type of test that recurs so frequently in the group tests as some form of this. Some form or forms of it are to be found in the following tests:—the Alpha Army test, the Beta Army test, the Trabue language test, and group tests devised by Whipple, Otis, Ballard, Haggerty and Thorndike. Some of the tests are conducted by giving sentences from which words have been omitted and the subject has to complete the sentence by the insertion of some word which makes sense. In other cases the given datum is a picture from which some feature is missing and the child is asked to supply the omission, i.e., to complete the picture. You will recall a simple form of the test in the Binet series where several faces are given from each of which something is missing—the nose, an eye, an ear, or the mouth. Obviously this test is one the difficulty of which is capable of immense gradation, but in all cases the mental processes involved are of the same general type.

The test of following simple directions occurs in the Alpha, the Columbian, the Civil Service, the Otis, Haggerty, Thorndike and Trabue group tests. One form of the test is to present various geometrical figures with the directions to make different marks in certain of the figures. Another form is to present a variety of letters, figures or words with directions to make a variety of marks, like a circle around one, a check mark above another, to underline

[1] Vide, pp. 30, 49, 96.

another, and so on. Another form of the test is to call on the subject to make some simple logical judgement on the basis of given data and to record the judgement in accordance with specific directions. A Trabue test of that type is one in which the following four words are presented:

QUART BUSHEL PECK PINT

The problem is: "If a peck is a greater magnitude than a bushel, cross out the word 'pint' unless a pint holds a smaller quantity than a quart, in which case draw a line under the first word after bushel." This test, in its more simple forms, involves the ability to comprehend simple directions and to carry them out. In its more complex forms such as the last example cited, it is combined with other features such as the ability to compare and make a logical selection.

The opposites test and the test of pointing out similarities both test forms of the associative process. Otis, Terman and Thorndike all use both forms. Trabue uses the opposites test. Woodworth and Hollingworth have both experimented with the test. In the Stanford-Binet tests we had examples of the similarities test in which the subject is asked to state the similarities in two or three things. Another test we observed called into play a little higher form of the same essential processes, viz., that of giving the differences in abstract terms, for that involved a comprehension at once of the similar and dissimilar elements. In the opposites test a list of words is given and the subject is instructed to write opposite each word a word which means the exact opposite. Another form of the test which Terman employed is to present a list of pairs of words some of which are synonyms and others antonyms. On a line with each pair are the words, "same" and "opposite," and the subject is directed to underline the word which expresses the relation between the two words. The Terman test comprises the following list of words:

SAMPLES	fall—drop	same—opposite	
	north—south	same—opposite	
1.	expel—retain	same—opposite	1
2.	comfort—console	same—opposite	2
3.	waste—conserve	same—opposite	3
4.	monotony—variety	same—opposite	4
5.	quell—subdue	same—opposite	5
6.	major—minor	same—opposite	6
7.	boldness—audacity	same—opposite	7
8.	exult—rejoice	same—opposite	8
9.	prohibit—allow	same—opposite	9
10.	debase—degrade	same—opposite	10
11.	recline—stand	same—opposite	11
12.	approve—veto	same—opposite	12

13. amateur—expert ... ... same—opposite 13
14. evade—shun ... ... same—opposite 14
15. tart—acid ... ... same—opposite 15
16. concede—deny ... ... same—opposite 16
17. tonic—stimulant ... ... same—opposite 17
18. incite—quell ... ... same—opposite 18
19. economy—frugality... ... ... same—opposite 19
20. rash—prudent ... ... same—opposite 20
21. obtuse—acute ... ... same—opposite 21
22. transient—permanent ... ... same—opposite 22
23. expel—eject ... ... same—opposite 23
24. hoax—deception ... ... same—opposite 24
25. docile—submissive ... ... same—opposite 25
26. wax—wane ... ... same—opposite 26
27. incite—instigate ... ... same—opposite 27
28. reverence—veneration ... ... same—opposite 28
29. asset—liability ... ... same—opposite 29
30. appease—placate ... ... same—opposite 30

An examination of the associative processes will make it clear that this test is one which calls them into play. We need scarcely be reminded that there are three types which have been traditionally accepted as the forms of association, association by contiguity, by similarity ahd by contrast. The latter two are correlative types, like the obverse and reverse sides of a coin Psychologically the processes involved are essentially the same. The superior intelligents are always richer in associations, and the mentally defective are always poor in associations. Whipple gives some valuable information in regard to the reliability of the test and its correlation to intelligence. The investigations bring to light such facts as that pedagogically retarded subjects are always below the average in the performance of this test, that the tests correlates at ·85 with the performance of all the tests combined, that it does not depend too much on schooling, that facility increases with practice, and that fatigue affects the process adversely.

The rearrangement of dissected sentences appeared as a test of twelve-year-old mentality in the Stanford-Binet scale. It has also a prominent place in the group tests, the Alpha Army, Columbian, Otis, Terman, Thorndike and Miller tests all including tests of this form. In several of the tests given the rearrangement of dissected sentences is combined with the true-and-false test. The form of the test is to present the words of a sentence in a disarranged order. The subject is asked to think what the sentence would assert were the words in correct order, and then to judge whether the statement made be true or false. Accordingly the words "true" and "false" are printed in a line with each sentence, and the subject has to underscore the word which signifies the truth or

falsity of the sentence when correctly arranged. The Terman test includes the following sentences:

SAMPLES	hear are with to ears	true—false	
	eat gunpowder to good is	true—false	
1.	true bought cannot friendship be	true—false	1
2.	good sea drink to is water	true—false	2
3.	of is the peace war opposite	true—false	3
4.	get grow they as children taller older ...	true—false	4
5.	horses automobile an are than slower ...	true—false	5
6.	never deeds rewarded be should good ...	true—false	6
7.	four hundred all pages contain books ...	true—false	7
8.	to advice sometimes is good follow hard ...	true—false	8
9.	envy bad greed traits are and	true—false	9
10.	grow an than peas palm tree higher ...	true—false	10
11.	external deceive never appearances us ...	true—false	11
12.	never is man what show a deeds	true—false	12
13.	hatred bad unfriendliness traits are and ...	true—false	13
14.	often judge can we actions man his by a ...	true—false	14
15.	in are always American cities born presidents	true—false	15
16.	certain always death of cause kinds sickness	true—false	16
17.	are sheet blankets as as a never warm ...	true—false	17
18.	never who heedless those stumble are ...	true—false	18

The value of the dissected sentence to be rearranged as a psychological test was discussed in connection with the twelve-year tests. The new factor here is the introduction of the true and false alternative upon which the subject has to make decision. This is a form of test which has come in for a good deal of criticism on the ground that a subject has a chance of guessing a fair proportion of the answers correctly. On the face of it that appears to be true, but it does not take into account all of the factors. In his recent book, *How to Measure in Education*,[1] Professor Wm. A. McCall of Columbia University has a good discussion of the utility and reliability of this test. He first of all points out its usefulness for an informational test, giving a sample test in the Geography of the United States. Twenty questions are asked. In the sample answer reproduced the subject had fourteen correct, five incorrect and one omission. His score would be 14 minus 5. The reason that his incorrect answers are deducted from the correct ones is to make allowance for this element of chance. McCall says: "Imagine a pupil who is absolutely innocent of any knowledge of the physical features of the United States." Were such a pupil to take the above test and were he to mark every statement he would, according to the theory of chance, mark ten statements correctly and ten incorrectly.

[1] New York: Macmillan, 1922. Vide pp. 119 ff.

The chances of guessing right or wrong are fifty-fifty or one to one. His score on the above test would be:

Score = 10 − 10 = 0.

In short, the pupil's knowledge is zero and the method of computing the score gives him zero. Suppose instead that he knows ten statements and guesses at the other ten. Of the ten guessed at he would, according to chance, get five correct and five wrong. That is, even though his real knowledge is ten he will show fifteen correct (10 + 5) and five incorrect. The method of computing his score brings out his real knowledge.

Score = 15 − 5 = 10.

A pupil who marks every statement correctly makes a perfect score, viz.:

Score = 20 − 0 = 20."

McCall points out that this method of scoring is used where the time allowed is brief, but where a great deal of time is allowed so that even the slowest pupils can complete the test it is customary to deduct double the number of incorrect answers. In the former case no account is taken of omissions; in the latter case there will scarcely ever be any omissions because the pupil is encouraged to guess at those which he does not know. A great many people insist that this test will always be to the advantage of the luckiest guesser. But the mathematical operations of the law of chance are inclined to refute that objection. In the long run "chance is fatally exact," so that the opportunities for injustice by this method of scoring are not great, especially where there are a large number of answers to be given on this plan. The test has this obvious advantage that it permits the examiner to get a good deal of information about the knowledge that a pupil possesses in a short time, without the necessity of reading examination papers in which other equally harmful elements have an opportunity of arising. The examiner who makes use of the true-false method needs to bear in mind several factors, because unless the test is carried out with rigid care, the opportunities for the miscarriage of justice are disproportionately large. He needs to bear in mind (1) the advisability of so devising the test that it will call for approximately the same number of responses of the two types; (2) the necessity of avoiding all ambiguity in the wording of his statements; (3) the necessity of critically examining the test which he is using to ascertain exactly what it measures; (4) the necessity of arduously avoiding suggestions as to the right or wrong answers; (5) the advisability of using no negative statements; (6) the wisdom of making the statements all concise; (7) the need of carefully avoiding too intimate connections between the succeeding statements so that any one will give a hint to the answer of another; and (8) the need to studiously avoid using the test for trivial statements.

I shall not take the time to discuss the various types of arithmetical tests which are in use in the various group tests of intelligence, but shall give some attention to them in the chapter on Tests of Attainment. Suffice it to point out that arithmetic as a whole involves many psychological processes and that arithmetical tests are not merely tests of schooling but that successful performance necessitates the functioning of intelligent processes. As a sample of the type of test I may give, however, the Terman test of arithmetic which I have altered so as to make it intelligible to Indian pupils.

1. How many hours will it take a person to go 66 miles at the rate of 6 miles an hour? *Answer*
2. At the rate of 2 for 4 annas how many pencils can you buy for Rs. 3? *Answer*
3. If a man earns Rs. 20 a month and spends Rs. 14 how long will it take him to save Rs. 300? *Answer*
4. $2\times3\times4\times6$ is how many times as much as 3×4? *Answer*
5. If two cakes cost Rs. 4-2-0 what does a sixth of a cake cost? *Answer*
6. What is $16\frac{2}{3}$ per cent of Rs. 120 *Answer*
7. 4 per cent of Rs. 1,000 is the same as 8 per cent of what amount? *Answer*
8. A has Rs. 180, B has $\frac{2}{3}$ as much as A, and C has $\frac{1}{2}$ as much as B. How much have all together? *Answer*
9. The capacity of a rectangular bin is 48 cubic feet. If the bin is 6 feet long and 4 feet wide, how deep is it? *Answer*
10. If it takes 7 men 2 days to dig a 140-foot ditch. how many men are needed to dig it in half a day? *Answer*
11. A man spends $\frac{1}{4}$ of his salary for board and room, and $\frac{3}{8}$ for all other expenses. What per cent of his salary does he save? *Answer*
12. If a man runs 100 yards in 10 seconds, how many feet does he run in 1/5 of a second? *Answer*

The analogies test is another example of a test of controlled association. The tests finds a place in the following Group Tests —the Columbian, Alpha, and those of Terman, Thorndike, Miller, Otis, and has been experimented upon by Burt and Woodworth. The usual form of the test is to state a relationship between two objects, give a third object and ask the subject to select from a list that is given the appropriate one to complete the analogy. An example would be that scissors are to paper as saw is to wood. In

this case the first part of the statement would be made, and the word *saw* given, after which a list of words such as *table, wood, shrub* and *tree.* The subject would be asked to underline the appropriate word to make the analogy plain. In the Thorndike test for High School Graduates which is used as part of the entrance examination to Columbia University the analogy test is given in pictorial form which makes it a little more difficult. One example will suffice to show how it is conducted. The second line contains the following pictures :—a comb—a whisp of hair—a tooth-brush—some teeth—an eye—a hair brush—the back of a bald-headed man's head. The test calls for a circle to be drawn around the object which bears the same relation to the third object which the second one does to the first. The following test used in the Terman group will serve as an example of the manner in which the analogy test is ordinarily employed.

(1) Coat is to wear as bread is to
eat starve water cook 1
(2) Week is to month as month is to
year hour minute century ... 2
(3) Monday is to Tuesday as Friday is to
week Thursday day Saturday ... 3
(4) Tell is to told as speak is to
sing spoke speaking sang ... 4
(5) Lion is to animal as rose is to
smell leaf plant thorn ... 5
(6) Cat is to tiger as dog is to
wolf bark bite snap 6
(7) Success is to joy as failure is to
sadness luck fail work ... 7
(8) Liberty is to freedom as bondage is to
negro slavery free suffer ... 8
(9) Cry is to laugh as sadness is to
death joy coffin doctor ... 9
(10) Tiger is to hair as mahseer is to
water fish scales swims ... 10
(11) 1 is to 3 as 9 is to
18 27 36 45 11
(12) Lead is to heavy as cotton is to
bottle weight light float ... 12
(13) Poison is to death as food is to
eat bird life bad 13
(14) 4 is to 16 as 5 is to
7 45 35 25 14
(15) Food is to hunger as water is to
drink clear thirst pure... ... 15
(16) b is to d as second is to
third later fourth last 16

(17) Village is to headman as army is to
navy soldier general private ... 17
(18) Here is to there as this is to
these those that then 18
(19) Subject is to predicate as noun is to
pronoun adverb verb adjective 19
(20) Corrupt is to depraved as sacred is to
Bible hallowed prayer Sunday 20

Investigation shows that the analogies test affords a high degree of correlation with general intelligence. Whipple quotes Wyatt's findings that the correlation was the next highest to the completion test of all. Burt also testifies to its reliability. According to Whipple, it " appears to be better suited than other tests of association to bring out individual difference in quickness of adaptation to the task demanded."[1]

The absurdities test was discussed in connection with the individual tests for ten-year-old intelligence.[2] There it was observed that the test had been found to measure intelligence very well when it was used in a fool-proof form. When Binet first devised the test he neglected to put it in that form. He did not tell his hearers that there was any absurdity contained in the statement which he was making with the result that he was greeted with shouts of ironical laughter. But when they were informed beforehand that he was about to read a statement which contained an absurdity which they would be asked to identify, they tackled the problem seriously. Dr. Ballard embodied the absurdities test in the Chelsea Mental Tests at the same time making them fool-proof. He gives 25 statements each of which contains something silly; and after each statement there are four attempts to point out what is foolish in it. The subject is required to read the statement and the four answers, and to point out which of the four he considers to be the most satisfactory. The following sample is given: " A soldier writing home to his mother said: 'I am writing this letter with a sword in one hand and a pistol in the other.' Foolish because—

A. The pistol might go off.
B. He could not write with a sword.
C. He could not write with both hands occupied.
D. Perhaps his mother could not read.

The best reason is the third; therefore C should be put on the answer paper." He is then asked to work the 25 statements in the same way. I may add two or three samples to illustrate:—

" 3. An old gentleman complained that he could no longer walk round the park as he used to: he could now only go half-way round and back again. Foolish because—

K. It would be better to walk into the country.

---

[1] Manual, Vol. II, p. 93. [2] Vide pp. 59 f.

L. The distances are the same.
M. He was getting lazy.
N. All old people are infirm."

"9. The moon is more useful than the sun, for it gives us light in the night when we really need it, while the sun gives us light in the day when we don't need it. Foolish because--

K. When there is a moon the night is not dark.
L. The moon is not so bright as the sun.
M. On some nights there is no moon at all.
N. It is the sun that makes the day."

"18. I am not conceited, for I don't think I am half as clever as I really am. Foolish because—

W. He is not so clever as he thinks he is.
X. He says he thinks he is clever and not clever at the same time.
Y. He can be clever without being conceited.
Z. A man should not brag about his cleverness."

In the first form in which Ballard used the absurdities test there were 34 statements of absurdities with which were mixed 4 quasi absurdities, and there were no suggested solutions, but the subjects had to supply them, as well as to point out the quasi absurdities. The form given in the Chelsea Mental Tests has been found to be fool-proof whereas the other was not so well-arranged. Yet even in its original form it was found to be a very valuable measure. The percentage of absurdities detected and explained was standardized as a measurement of the grade of intelligence. Ballard's standardization gives the following results :—

Average:	13·1	14·4	15·1	17·4	18·5	18·9	18·9
Age:	11	12	13	14	15	16	17

Another type of test which is being used by Terman and Thomson is the classification test, the latter including it in the Northumberland Mental Tests of which he is the author. The form in which he uses the test is five rows of words; four of which belong to a group and one of which is not homogeneous, with the instruction that the extra word is to be crossed out in each case. The following is the form of the test :—

charity	kindness	benevolence	revenge	love
square	circular	oblong	hexagonal	triangular
needle	tack	nail	knife	pin
coal	bread	coke	wood	paper
bran	wool	cotton	hemp	jute
hair	feathers	wool	grass	fur

I append also a copy of the Terman classification test in a form of adaptation which seems to me to make it more suitable to India.

SAMPLES { 1. ~~bullet~~ cannon gun sword
2. England ~~Bombay~~ China India America

In each line cross out the word that does not belong there. Cross out JUST ONE WORD in each line.

1. Moses Raman Gopal Venkuswami Ratnam.
2. Brahman Panchama Sudra Reddy Englishman.
3. automobile bicycle ox-cart house train.
4. ox calf ram cow bull.
5. hop run crawl stand walk.
6. death grief picnic poverty sadness.
7. bed chair vessel bench table.
8. hard rough smooth soft sweet.
9. cooly doctor lawyer priest teacher.
10. Jesus Buddha Muhammad Krishna Gokhale.
11. butterfly hawk crow parrot myna.
12. cloth cotton flax hemp wool.
13. digestion hearing sight smell touch.
14. down hither recent up yonder.
15. anger hatred joy pity reasoning.
16. Australia Cuba Andaman Ceylon Burma.
17. Arjuna Krishna Clive Kali Hanuman.
18. give lend lose keep waste.

Another form of the classification test which is included in the Northumberland tests is one in which the subject is required to cross out extra numbers from five lines of numbers. The mental processes involved in classification are the higher processes in which analysis and synthesis take part. It involves the ability to form concepts and to make judgements on the basis of the concepts formed. This is a logical process, and success involves the forming of a class concept and the comparison of the individual data with the generalization to ascertain what may be subsumed and what not. This is the type of reasoning which enters into all of our higher mental processes and therefore such a test calls for a response which only an intelligent person can make.

The test of learning to use a code was mentioned as one of the tests of average adult intelligence. There are tests which are not quite the same, yet call for similar mental processes among the group tests. For example, the Chelsea Mental Tests include a ciphering test in which certain signs which we usually use as punctuation marks are used in the place of the vowels and letter *h* wherever these occur in words. Then questions are asked which call for the re-translation of the ciphers into the letters which they symbolize before intelligent answers can be given. The problem is put in the form of a necessary device by informing the subjects that the printer once lost all of his type for the vowels and the letter *h* and had to substitute punctuation marks which may be interpreted in accordance with the following key:—

	a	e	i	o	u	h
KEY	,	.	;	:	!	-

Then the subject is informed that there are twenty-five questions printed in this funny way which he must decipher and answer. A sample sentence is given which the instruction sheet interprets, and then the subjects are set at the problems. The cipher test is one with which the examiners require to take especial care. Care must be taken to be sure that the subjects comprehend what is required, but at the same time they should not be given a chance to memorize the key. On that account the preliminary arrangements are fixed and standardized, three minutes being allowed for the reading of the instructions, and ten minutes for the performance of the test. The performance calls into play the ability to form new associations. Language is a type of symbolism, and letters are signs for sounds. We have been habituated to the use of certain signs for sounds and words, and the test calls for the substitution of a new set of symbols to displace some of the old. The learning process is called into play, and we are made to realize the grip of habits through the difficulty which the contrast forces upon us.

Certain tests have been included by some psychologists which are designed to measure the correctness of the logical processes in memory and selection. Whipple's account of the work done on his test is not quite reassuring. The logical memory test is intended to discover the ability of the subject to remember and reproduce ideas in a logical order, and differs from the rote memory test where a reproduction of words remembered is considered as satisfactory. It is better calculated than the rote memory test to discover individual differences in memory efficiency, and from that point of view the reliability of the test is acceptable. But as a test of intelligence it correlates rather lower than would be anticipated. However when sub-normals were tested with this test the result of their responses was in close accord with their general mentality as tested by the Binet method.

The logical selection tests are somewhat different. In these instances reproduction from memory is not demanded, and memory comes into play only as affording from the experiences of the past the information which will enable the subject to give the correct response. Indeed some examiners, as Trabue for instance, have included this test under the head of "information tests" rather than captioning it as a test of logical selection which Terman has done. Here the usual procedure is to give a sentence, all but the last word, and a selection of words from which the subject is to mark the one which completes the sentence. It will be seen that the completion method is involved here also, so that what has been said about the processes called forth by the completion tests

applies equally here. The Trabue form of the test may be illustrated by two or three examples, as follows:—

A SAW is used by a	○ PAINTER	○ PLUMBER	○ CARPENTER	○ MASON
CORAL is found in	○ TREES	○ REEFS	○ MOLLUSKS	○ MINES
AN EMERALD is	○ GREEN	○ RED	○ BLUE	○ BLACK

The subject is expected to insert a check mark (✓) in the circle in front of that one of the four words which makes the best sentence and tells the most exact truth. In the Terman form of the test the sentence may be completed by two words, so that the subject, instead of selecting one out of four words, selects two out of five which are to be underlined. The test is as follows:—

SAMPLE.—A man always has
body cap gloves mouth money.

1 A horse always has
harness, hoofs shoes stable tail. 1

2 A circle always has
altitude circumference latitude longitude radius. 2

3 A bird always has
bones eggs beak nest song. 3

4 Music always has
listener piano rhythm sound violin. 4

5 An object always has
smell size taste value weight. 5

6 Conversation always has
agreement persons questions wit speech. 6

7 A banquet always has
food music persons speeches toastmaster. 7

8 A pistol always has
barrel bullet cartridge sights trigger. 8

9 A ship always has
engine guns keel rudder sails. 9

10 A debt always involves
creditor debtor interest mortgage payment. 10

11 A game always has
cards contestants forfeits penalties rules. 11

12 A magazine always has
advertisements paper pictures print stories. 12

13 A museum always has
animals arrangement collections minerals visitors. 13

14 A forest always has
animals flowers shade underbrush trees. 14

15 A citizen always has
country occupation privileges property vote. 15

16 A controversy always involves
claims disagreement dislike enmity hatred. 16

17 War always has
airplanes cannons combat rifles soldiers. 17

18 Obstacles always bring
difficulty discouragement failure hindrance stimulation. 18

19 Abhorrence always involves
aversion dislike fear rage timidity. 19

20 Compromise always involves
adjustment agreement friendship respect satisfaction. 20

Tests of practical judgement are used by Haggerty, Terman and Thorndike and also in the Columbian and the Alpha Army tests. The tests are tests of common sense. In the Alpha test sixteen simple questions are asked, and below each question three answers are given. The subject is asked to examine the answers carefully and place a cross in the circle opposite the one which makes the best answer. One or two examples will illustrate the type :—

" Why are pencils more commonly carried than fountain pens? Because—

O they are highly coloured.
O they are cheaper.
O they are not so heavy.

" Why is leather used for shoes? Because
O it is produced in all countries.
O it wears well.
O it is an animal product."

Terman designates the test as " best answer " test. The following is the test as he devised it, adapted for India :—

SAMPLE
Why do we buy clocks? Because
1. We like to hear them strike.
2. They have hands.
× 3. They tell us the time.

1. Spokes of a wheel are often made of junglewood, because
1. Junglewood is tough.
2. It cuts easily.
3. It takes paint nicely.

2. The saying " A watched pot never boils," means
1. We should never watch a pot on the fire.
2. Boiling takes a long time.
3. Time passes slowly when we are waiting for something.

3. A train is harder to stop than an automobile, because
1. It has more wheels.
2. It is heavier.
3. Its brakes are not so good.

4. The saying "Make hay while the sun shines," means
   1. Hay is made in summer.
   2. We should make the most of our opportunities.
   3. Hay should not be cut at night.

5. If the earth were nearer the sun
   1. The stars would disappear.
   2. Our months would be longer.
   3. The earth would be warmer.

6. The saying "If wishes were horses, beggars would ride" means
   1. Wishing does n't get us very far.
   2. Beggars often wish for horses to ride.
   3. Beggars are always asking for something.

7. The saying, "Continual dropping wears away a stone," means
   1. Stone is not strong.
   2. Continual dropping is not a good thing.
   3. Continued effort brings results.

8. A kite flies, because
   1. It has a tail.
   2. It is made of light material.
   3. It has bright colours.

9. The feathers on a bird's wings help him to fly, because
   1. They make a wide, light surface.
   2. They keep the air off his body.
   3. They decrease the bird's weight.

10. The saying "A carpenter should stick to his bench" means
   1. Carpenters should not work without benches.
   2. Carpenters should not be idle.
   3. One should work at the thing he can do best.

11. The saying "If the rider is lenient the horse goes on three legs" means
   1. If the overseer is too lenient the coolies will be lazy.
   2. Horses do not like easy riders.
   3. Horses walk on three legs.

Terman and Thorndike both use tests which involve carrying on a number series from a given amount of data. The Northumberland and Columbian also include tests of this type. In each case the subject is advised to study the series as far as given so as to discover the principle of progression, and then to carry on the series to one or two more places for which blank spaces are provided. The test is of one's ability to comprehend quickly and accurately the relations between series of numbers. The Thorndike and Terman series are very similar, except that the former calls for only one additional place to be filled in, while the latter calls for two.

The Thorndike test, we recall, is for High School graduates and as might be expected includes one or two series slightly more difficult than the Terman series, as, e.g., one series which progresses by the addition of 5/12 to each previous result. The following is the Terman test:—

SAMPLES	5	10	15	20	25	30	35
	20	18	16	14	12	10	8

In each row try to find out how the numbers are made up, then on the two dotted lines write the TWO numbers that should come next.

1st row			8	7	6	5	4	3	....
2nd row		3	8	13	18	23	28	....	
3rd row	11¾	12	12¼	12½	12¾	....			
4th row			8	8	6	6	4	4	....
5th row			1	2	4	8	16	32	....
6th row		4	3	5	4	6	5	7	....
7th row			16	8	4	2	1	½	....
8th row			8	9	12	13	16	17	....
9th row	7	11	15	16	20	24	25	29	....
10th row	31·3	40·3	49·3	58·3	67·3	76·3	....		
11th row					1/20	1/4	1	5	....
12th row			3	4	6	9	13	18	....

The methods of scoring for the various group tests are not so different as in the case of the individual tests. Practically all of the examiners who have devised tests use a system of credits. In each case a scale of instructions for scoring accompanies the tests, and those who use them must follow these instructions to gain the most from the use of the tests. For the psychologist who arranged the scale has also standardized the results so that one can judge the mentality of subjects on the basis of results. Tables of equivalent ratings in various scales have been worked out and may be consulted when desired. Yoakum and Yerkes in *Army Mental Tests* (p. 133) gives such a table for the Alpha, Beta, Point Scale, Performance and Stanford-Binet scales. Wilson and Hoke in *How to Measure* (p. 251) give a similar table for the Trabue Language scale and the Binet-Simon scale.

There has been a very keen discussion among psychologists in regard to the relative value of the individual and group tests. The protagonists of the individual tests have argued that they afford a more accurate diagnosis of the individual tested than the group test can hope to do. Whipple voices that criticism in the statement that "on the whole, and especially when careful analytic work is contemplated, the group method, save for the preliminary trial of a method, is out of place. There are almost sure to be some subjects

in every group that, for one reason or another, fail to follow instructions or to execute the test to the best of their ability. The individual method allows the examiner to detect these cases, and in general by the exercise of personal supervision, to gain valuable information concerning the subject's attitude towards the test." [1]

On the other hand the group tests certainly have the best of the argument in the matter of the time economized. Under certain circumstances group testing is also found to be fairer to those tested. The personal equation of the examiner is much less likely to enter such calculations, and standardization is more readily effected. I may conclude with the words of McCall:

" What then is the conclusion of the whole matter? Individual testing and group testing each secure special values. The method adopted in the psychological examination of soldiers will probably come into common use in all educational measurements, whether done for purely pedagogical or for clinical purposes. The initial tests given to the soldiers were group tests. They revealed the illiterates and those that were in some way abnormal. The illiterate and abnormal groups were then intensively measured by individual tests. The diagnoses afforded by the group tests were accepted for the vast majority of the recruits. In time school psychologists will not wait until abnormal cases are sent to them for diagnosis. They will sweep through the schools with a net of group tests and catch their own cases for intensive study. Even for the special cases, what with the development of group tests for illiterates, it is worth considering whether the greater number of group tests which may be given within an equal time-interval may not give a better diagnosis than a fewer individual tests. A good practical rule is to first give group tests, accept their diagnosis for most of the pupils, and give further group or individual tests to the few pupils, who, according to the group tests, need special study."[2]

---

[1] Manual, Vol. 1, p. 8.

[2] How to Measure in Education, p. 235.

CHAPTER VII.

VOCATIONAL TESTS AND TESTS OF CHARACTER.

A.—VOCATIONAL TESTS.

A specialized use of the psychological test is in the detection of vocational fitness. The method of procedure is twofold. One is to use the intelligence tests as a measurement of vocational ability: the other is to use tests that have been specially devised to detect vocational fitness of a particular type. In the first chapter we noted some of the earlier and less scientific attempts to determine mental characteristics, including phrenology and physiognomy. In both of these cases the results were regarded as useful in vocational diagnosis. But with the passing of the old structural manner of classifying mental processes, it became apparent that such methods could lay no claim to reliability. Attending, cognizing, feeling, willing, remembering, reasoning, judging, perceiving, and all of the other mental phenomena that were at one stage in the history of psychology treated as faculties or powers are now universally regarded as processes or functions.

The most interesting, perhaps because the most consistent, of the faculty psychologists was Rudolph Herman Lotze (1817—1881). Baldwin in his *History of Psychology* sums up Lotze's position as follows:

"Put on the defensive in the matter of determining the fundamental functions or faculties, Lotze accepted the consequences of his view. Herbart and Brentano had argued that if once we admit faculties, there is no stopping anywhere; every distinguishable mode of mental process may be described as a separate faculty; colour-perception and piano-playing no less than feeling and will. Lotze did not deny this, but claimed that certain generalizations were possible which permitted the demarcation of the great functions recognized in the Kantian threefold division."[1]

The change from that point of view is complete. When we desire to discover ability in playing the piano or in any other art, profession, or other calling, we no longer expect to account for it in terms of any individual faculty, nor do we search for it as though it were something distinct and distinguishable. That piecemeal fashion of dealing with mental processes has ceased, because the progress of psychology has established the fundamental unity of the mental processes. And the logic of unity is expressed in complexity. If the processes intertwine and interlink, the

[1] Vol. II, p. 34.

difficulty of dealing with any one of them disparately is evident. When the person or subject is cognitive it means that the whole person or subject is at that time engaged in the experience of cognizing and not that one portion of the mind is doing that while the other portions are going on with their work undisturbed. And what has been remarked about cognition applies with equal force to all the mental processes. The whole person knows, feels, chooses, remembers, imagines, or engages in whatever the mental process it may be that is under discussion.

The meaning of this for vocational psychology should be fairly obvious. It signifies that when any one plays a piano, or drives a motor-car, or weaves a cloth, or cobbles a shoe, or moulds a pot, or engages in any other task, that it is the person who is so engaged, and not merely some power of faculty that is kept employed while the others lie dormant. The processes which are involved in any occupation, be it never so simple, are much more complex than it was supposed. So the vocational test will have to take account of the complexity of the process in trying to measure fitness for any specific calling.

The interest in vocational tests has been inspired from two sources. In the first place there have been industrial services which have turned to the psychologist to assist them in the task of selecting men for the recruitment of their services. In the second place there is the newly developed educational interest in vocational guidance. It is the task of education to do all that is within the range of possibility to prepare a person for complete conscious adjustment to his environing world. That involves a consideration of the best way that a person can contribute to the community's welfare, the best form of service which he is fitted to render—a vocational consideration.

There are a number of contributive factors that enter into the matter of vocational guidance, and which must be considered by the psychologist who is undertaking to provide tests for selective purposes. These have been admirably summarized by McCall as follows:

"(1) A careful survey of the various occupations to determine the constancy of demand for employees, whether the occupation is a seasonal or ephemeral one, the ratio of demand to supply, the monetary rewards, the nature and amount of other types of rewards, the working conditions in the occupation, etc.; (2) a study of the results of such a survey by the pupil, both to aid him to choose his own occupation intelligently and as an important part of his general education; (3) a testing of various ways of the pupil's ability for and interest in each of the occupations; (4) the choice by the pupil with the advice of a vocational counsellor of his vocation; (5) the provision of adequate vocational education; (6) appropriate educational guidance in the light of the chosen

vocation; (7) vocational placement at the end of the pupils educational preparation; and (8) a systematic follow-up of each pupil sent to industry." [1]

The function of the vocational test may be understood in relation to the whole process by such a survey as that quoted. Its aim is strictly practical—to serve as an aid in vocational selection, and after the selection has been determined ought to be followed by vocational education adequate to the demands of the case.

Reference has already been made to the fact that vocational fitness has been tested sometimes by means of intelligence tests. The reason for this is not hard to seek. There are some occupations which call for higher mental processes for their successful performance than others. To be sure, we expect a man of superior mental ability to perform his work, whatever its nature, better than another man of inferior ability. But there are other tasks which demand the functioning of such complex processes that only persons of high levels of intelligence are capable of succeeding in them. Quite obviously it requires more intelligence to serve efficiently in the Legislative Council than it does to perform the duties of a gardener. The school teacher is expected to be a more intelligent person than the cooly. Among the backward classes there may be some who could have been fitted for school teaching and membership in the Legislature if they had been more fortunate in regard to schooling and other social opportunities. But in a community where there is a democracy of privilege, we expect to find certain occupations occupied by the more intelligent.

The Army Mental Tests brought to light considerable information in regard to the relation between vocational fitness and intelligence, information which must be of value to the vocational educator. It will be recalled that over 1,700,000 men were examined by the Army psychologists. That means that a great deal of data was assembled and has been made available in regard to various phases of the subject. Since the records show the occupation of every enlisted man who was examined, it has been possible to classify the men by occupations and to record the results of their intelligence examination, and, by taking the averages, to determine what is the average intelligence of the men coming from the various occupations. The War Department issued a bulletin on the subject with the following table of average scores:—

120 and over—Army chaplains and engineer officers.

115—119—Stenographers, typists, accountants, civil engineers, Y.M.C.A. secretaries and medical officers.

110—114—Mechanical draughtsmen.

105—109—Mechanical engineers.

[1] How to Measure in Education, p. 170.

100--104—Book-keepers.

95—99--General clerks and filing clerks.

90--94—Railroad clerks.

85—89—Photographers.

80--84—General electricians, telegraphers, band musicians, and concrete construction foremen.

75—79--Receiving clerks, shipping clerks, and stock-keepers.

70--74—Truckmasters, farriers and veterinarians.

65--69—Laundrymen, plumbers, auto repairers, general pipe-fitters, auto engine mechanics, auto assemblers, general mechanics, tool and gauge makers, stock checkers, detectives and policemen, toolroom experts, ship carpenters, gunsmiths, marine engineers, hand riveters, and telephone operators.

60--64—General machinists, lathe hands, general blacksmiths, brakemen, locomotive firemen, auto chauffeurs, telegraph and telephone linemen, butchers, bridge carpenters, railroad guards, railroad shop mechanics, and locomotive engineers.

55—59—General carpenters, painters, heavy truck chauffeurs, horse trainers, bakers, cooks, concrete and cement workers, mine drill runners, bricklayers, cobblers and caterers.

50—54—Stationary gas engine men, horse hostlers, horse shoers, tailors, general boilermakers, and barbers.

45—49—Farmers, labourers, general miners and teamsters.

The Army psychologists not only gathered data in regard to the previous occupations of the enlisted men, but they used the information which they obtained for vocational purposes. On the basis of the intelligence tests they recommended to the War Department as unfit for the vocation of a soldier in any capacity and therefore to be discharged 7,800 men. They recommended for service in labour battalions or other service battalions but of insufficient mentality for active service 10,014 men, and for further observation to be placed in development battalions another 9,487 men. Altogether they discovered 45,653 men of intelligence below the standard of ten years. Of these Yoakum and Yerkes say: "It is extremely improbable that many of these individuals were worth what it cost the government to maintain, equip, and train them for military service." [1]

Mention has already been made of the fact that the system of scoring adapted was a letter gradation beginning with "A" and ending with "E." In interpreting the meaning of these letters, vocational fitness was evidently one of the guiding principles. I quote from Yoakum and Yerkes some of the relevant items:

A = Very superior intelligence . . . "A" men are of high officer type when they are also endowed with leadership and other necessary qualities.

---

[1] Army Mental Tests, p. 21.

B = Superior intelligence. . . . The group contains many men of the commissioned officer type and a large amount of non-commissioned officer material.

C + = High average intelligence. This group . . . contains a large amount of non-commissioned officer material with occasionally a man whose leadership and power to command fit him for commissioned rank.

C = Average intelligence. . . . Excellent private type with a certain amount of fair non-commissioned officer material.

C— = Low average intelligence . . . "C—" men are usually good privates and satisfactory in work of a routine nature.

D = Inferior intelligence . . . "D" men are likely to be fair soldiers but they . . . rarely go above the rank of private.

D— and E = Very inferior intelligence . . . (1) "D—" men are considered fit for regular service; and (2) "E" men, those whose mental inferiority justifies their recommendation for development battalion, special service organization, rejection, or discharge.[1]

In one sense the whole mechanism of the Army Mental Tests was evolved for vocational ends. It was to discover the best way of organizing the material available so as to produce an efficient army in the shortest possible time that the psychologists were mobilized. So that the whole experiment is one of immense importance for the subject under consideration. The tests recognized that there was one element essential to the equipment of a good soldier, viz., intelligence, an element that was measurable. Not only so, but it was recognized that there is no other single factor commensurate with intelligence for the soldier's equipment. They were not expected to measure loyalty, endurance, courage and the ability to command, but it was discovered before they finished that such qualities were much more frequently present in men of superior intelligence than in any other group of men. A ruling was made, after a certain amount of experience had been gained that no man should be accepted for an Officers' Training School whose score was below the "C+" grade, unless he showed most extraordinary ability in other directions. They also found it inadvisable, unless the circumstances were very exceptional, to accept men below the "C" rank of intelligence for the posts of non-commissioned officers. Men below the standard of "C" were found to be scarcely ever capable of doing complicated clerical work. Certain branches of the service, such as Signal Corps, Machine Gun Operators, Field Artillery and Engineers, were found to require men of superior intelligence, and were organized with twice the proportion of superior men as the ordinary branches of the service.

[1] Yoakum and Yerkes : Op. cit., pp. 22, 23.

After pointing out these results, it can scarcely need emphasis that the results are general. Within the limits of various occupations will be found great variations, and the results recorded are the average for a large number. The upper and lower limits are both significant. The upper limit of a vocation indicates the point beyond which subjects cease to have any interest in such an avocation. The lower limit indicates the point below which the subject would not have sufficient intelligence to meet the demands of the occupation. There are two uses, then, for the intelligence test in this connection. The one is, the examination of large numbers employed within a vocation, to determine the limits within which one may succeed within the various occupations. The other is to measure the individual subjects to determine in what occupational levels his intelligence comes. To be sure there is a great deal of overlapping here. But the possibilities can be discovered by this method, and it can often be ascertained that certain occupations are impossible to a subject, though he or she may be seeking such employment because of the lack of other occupation.

Of course there are other factors which must not be neglected. We must not expect the intelligence test to determine everything that we need to know to fix a man's vocational fitness. Moral fitness is an important factor. Says Hollingworth: "What one lacks in quickness it is often possible to make up in persistence; what another lacks in ambition and competitiveness he may supply in the form of loyalty and zeal; relative intellectual inferiority is often and easily balanced by the display of the social charm; persistent, well-directed and enthusiastic effort or even a good vocabulary may enable one to compete successfully with the exceptional genius who does not display these incentives to advantage . . . I would rather trust my life and limb to a motorman whose feeble memory span is re-enforced by a loyal devotion to the comfort of his grandmother than to a mnemonic prodigy whose chief actuating motive in life is to be a 'good fellow'".[1]

Special aptitudes is a question that concerns the vocational psychologist. It happens at times that certain individuals have remarkably superior gifts for certain forms of occupation. And the reverse is true—some people of superior general intelligence are very inapt in certain particular occupations which call for special abilities. Carney describes[2] the case of a graduate of the

[1] Vocational Psychology, pp. 216, 217.

[2] Carney: Some Experiments with Mental Tests as an aid to selection and placement of clerical workers in a large factory; University of Indiana Bulletin, Vol. V, No , pp. 60—74.

University of Chicago, who was very keen intellectually and possessed with a charming personality, being employed in a large factory and given the task of computing percentages on a slide rule. To everybody's dismay, she was a pronounced failure. She was sent to Carney for a test who discovered that she was very high in intelligence but very low in arithmetical ability. She was changed to another department where general intelligence was needed rather than specialized mathematical ability, and in a short time rose by her superior intelligence to be the head of that department.

We need only revert to the matter discussed in the second chapter namely the complex character of intelligence to furnish evidence for the contention that when one is dealing with vocational aptitude, he is dealing with a problem which is sure to involve a great breadth of abilities. Some vocations demand keen mathematical ability; some ability in drawing; others quickness of visual perception; others specialized motor ability, and so on. If these specialized abilities are to be detected before the person is actually tried in an occupation, it means that specialized tests must be employed. In other words, tests of general intelligence are serviceable only in determining the limits within which various occupations fall, but do not discover any special aptitude which may be demanded by that particular occupation. The case of the young lady whom Carney describes is in point. That means that there is a place and a function for a specialized vocational test, in addition to the work of the intelligence test.

On the other hand there is a host of occupations which require no special aptitude of any kind, and these are the occupations which are filled by people of the ordinary or even inferior grades of mentality. A perusal of the results reached by the Army investigators will be sufficient to indicate that there are many occupations which are open to men and women of lower types of intelligence, and where honesty, truthfulness, patience, courtesy and such moral and social virtues are of more importance than special aptitudes of any type.

We shall now consider some of the specialized tests that have been employed for the detection of vocational fitness. One principle which we may find operating in many of the tests is that of creating a situation for the subject which shall have as many similar characteristics as possible to the occupation itself. Thus in connection with the selection of subjects to be recommended from commercial schools for clerical positions, it is common to assign them pieces of work similar to those for which the occupation calls, scoring their performances as successes or failures with reference to the occupations considered. Some of the forms of this test include the striking of a trial balance in bookkeeping, making certain commercial calculations, finding addresses, finding telephone numbers, carrying out verbal instructions, etc.

Sometimes subjects have been taken to a psychological laboratory where their performances can be observed closely by psychologists. Hollingworth cites as examples Thorndike's observations of candidates for clerical positions and positions of salesmen, Paynter's observations of candidates for the position of judges of trade-mark infringements, Scott's observations of salesmen, and others in the case of tests for handwriting experts.

Another type of test is that which seeks to create a situation which, while not exactly parallel to that of the occupation itself, attempts to test the functions and processes which the occupation calls forth by tests which involve similar attitudes and endeavours. That in itself is a matter which demands careful psychological analysis, for what appears on the surface to be the important function of the occupation does not always turn out that way in experimental observations. Münsterberg illustrated that in the case of type-setting. He recorded that his impression was that rapidity of performance depended upon the quickness of finger reaction. But the managers have observed, on the contrary, that the most essential condition for speed in the operation is the ability to retain a large number of words in memory before they are set, this ability more than counterbalancing any loss of speed in finger movement.[1] To select girls for positions as type-setters one of the tests employed has been speed of reaction to a sound stimulus.

Münsterberg conducted a series of experiments in an endeavour to devise tests for the selection of men for marine service. He was approached by one of the large ship companies to ascertain whether it were possible to devise a psychological test for ship officers, emphasis being placed upon the fact that such an officer must be one who can respond to an unexpected situation with quickness and accuracy. The company was well aware of the type of man needed and the types which would be dangerous. The type of man needed was one who could act appropriately when unexpectedly confronted with a complicated situation such as the speedy approach of another ship in a fog. There are two types which ought to be excluded. The one understands precisely what is required but is paralyzed when a dangerous condition suddenly confronts him, vascillating between possible actions until any action is too late to be of service. The other type realizes the need for rapid action to save the situation, but under the pressure of the danger involved acts with absolutely no deliberation, doing the first thing that suggests itself at the time.

Münsterberg realized that the complex type of reaction called for involved several mental abilities, including processes of discrimination, association, memory, perception and suggestion. The

[1] Psychology and Industrial Efficiency, p. 124.

most important factor was the securing of an appropriate decision in a sufficiently short time, so that a test must reproduce in experimental form such a situation. "It would seem necessary," he says, "to create a situation in which a number of quantitatively measurable factors were combined without any one of them forcing itself to consciousness as the most important. The subject to be experimented on has to decide as quickly as possible which of the factors is the relatively strongest one." [1]

The test devised was in the form of sorting cards into appropriate piles in accordance with given directions. Twenty-four cards of the same size as playing cards were arranged, so that on the upper half of each card there were four rows of twelve capital letters, namely A, E, O, and U in irregular repetition. "On 4 cards, one of these vowels appear 21 times and each of the three others 9 times; on 8 cards, one appears 18 times and every one of the three others 10 times; on 8 cards, one appears 15 times and each of the three others 11 times; and finally, on 4 cards one vowel appears 16 times, and each of the three others 8 times, and besides them 8 different consonants are mixed in. The person to be tested has to distribute these 24 cards as quickly as possible in 4 piles, in such a way that in the first pile are placed all the cards in which the letter A is most frequent, in the second those in which the letter E predominates, and so on. As a matter of course the result must never be secured by counting the letters. Any attempt to act against this prescription and secretly to begin counting would moreover delay the decision so long that the final result would be an unsatisfactory achievement anyhow. It would accordingly brings no advantage to the candidate." [2]

Münsterberg believed that the reactions of different subjects to the card sorting experiment were parallel to those of the person engaged in practical ship service. Some persons lose their heads completely and exhibit that sort of mental paralysis which prevents a man from arriving at a conclusion which will meet the demands of a situation and be satisfying to himself and others. "Some chance letters stand out and appear to them to be predominant, but in the next moment the attention is captured by some other letters which bring out the suggestion that they are in the majority and that they present the most important factor. The outcome is that inner state of indecision which becomes so fatal in practical life." There are others again who go at the task with a rush, sorting the cards very speedily under the impression that the first impulse is correct. But this type of subject makes many mistakes which might be avoided with some deliberation. "Any small group of letters which catches their eye makes on them, under the pressure

[1] Op. cit., pp. 86, 87.
[2] *Ibid.*, pp. 87, 88.

of their haste, such a strong impression that all the other letters are inhibited for the moment and the wrong decision is quickly made." A third type of subject performs with a fair amount of speed and yet with sufficient care to get a majority of correct responses. Accurate visual perception obviously enters into a successful response, and the person responds with the feeling that the exercise is itself interesting and stimulating to the mental processes. The experimenter took into account the time of performance, the number of mistakes and the character of the mistakes, for it will be apparent that all errors were not equally significant. Where the predominance of one letter is less marked the chance of making an error is much greater than where the predominance is more marked.

Münsterberg worked on the belief that the best vocational test was one that did not present a miniature model of the exact situation, but rather one which calls for the functioning of the same inner psychological process. This writer says: " A reduced copy of an external apparatus may arouse ideas, feelings and volitions which have little in common with the processes of actual life . . . Experiments with small models of the actual industrial mechanism are hardly appropriate for investigations in the field of economic psychology. The essential point for the psychological experiment is not the external similarity of the apparatus, but exclusively the inner similarity of the mental attitude. The more the external mechanism with which or on which action is carried out becomes schematized, the more the action itself will appear in its true character." [1]

Another test for which Münsterberg is responsible is one for the selection of motormen for tram-cars. The need for such a test was emphasized by a study of the causes of accidents in various cities of the United States. The American Association for Labour Legislation called a meeting of specialists in 1912 which was to consider the problems raised by these accidents, and this investigation took into account many factors which entered into the matter. Fatigue was one of the prominent factors which was recognized, but beyond that there was also recognized the mental make-up of motormen. Obviously the occupation is such that successful performance depends upon a number of factors including attention, visual perception, ability to resist distractions, and speed in discerning the possibilities involved when certain conditions present themselves. Much of the discussion in regard to the marine service holds good in regard to electric railway service also. The mental response demanded of a motorman at the wheel of an electric tram-car is not unlike that demanded of a captain at the bridge when some sudden and unforeseen emergency arises.

[1] *Ibid.*, pp. 67, 68.

Münsterberg, after a considerable study of the data at hand, came to his own conclusions as to the mental processes involved. He writes: "I found this to be a particularly complicated act of attention by which the manifoldness of objects, the pedestrians, the carriages and automobiles, are continuously observed with reference to their rapidity and direction in the quickly changing panorama of the street. Moving figures come from the right and from the left toward and across the track, and are embedded in a stream of men and vehicles which moves parallel to the track. In the face of such manifoldness there are men whose impulses are almost inhibited and who instinctively desire to wait for the movement of the nearest objects; they would evidently be unfit for the service, as they would drive the electric car far too slowly. There are others who, even with the car at high speed, can adjust themselves for a time to the complex situation, but whose attention soon lapses, and while they are fixating a rather distant carriage, may overlook a pedestrian who carelessly crosses the track immediately in front of their car. In short we have a great variety of mental types of this characteristic unified activity, which may be understood as a particular combination of attention and imagination." [1]

Having determined against the principle of testing by means of models, this investigator proceeded to devise a test for motormen that would test such psychological abilities as attention and imagination which he found to be needed in the actual situation. The arrangement on which he settled was in the form of a card nine half-inches broad and twenty-six half-inches long. Two heavy lines half an inch apart were drawn lengthwise through the centre of the card, thus leaving a space of four half-inches on either side. The entire card is divided into half-inch squares. The two heavy central lines represent a tram-car track on a street, on either side of which are four rows of squares filled in an irregular way with black and red figures of the first three digits. The digit, "1" represents a pedestrian who moves just one step, and the digit "2" represents a horse which moves twice as fast, while the digit "3" represents an automobile which moves three times as fast. The black digits represent men, horses and automobiles moving parallel to the track and which cannot cross the track and therefore can never constitute a danger. The red digits represent men, horses and automobiles moving from either side toward the track and hence constituting a danger. The dangerous situations are when the red digit 3 is three units from the track, or the red 2 is two units from the track, or the red 1 is one unit from the track. If either of these is more units away than as indicated it signifies that the man, horse or automobile would not reach the track until the car

[1] Op. cit., p. 66.

has passed, or if they are less it indicates that they will cross over before the car arrives. The test is for the subject under examination to indicate as rapidly as possible the danger points on the diagram, a task that is complicated because of the presence of the black digits which divert the attention and because of the red figures which are either too near or too far to constitute dangers. Twelve cards of this kind were used in the experiment, the cards being placed one on the other and each with a handle, all of them under a glass plate. This entire apparatus is placed in a black wooden box, completely covered by a belt made of heavy black velvet which moves over two cylinders at the front and rear ends of the apparatus. In this belt are windows which move over the card with its track and figures. As the belt is revolved the subject under test has to call out the dangerous places, this being done for the twelve cards in succession while the experimenter times the performance and records the responses. The experiment is scored in accordance with three factors: the number of seconds occupied by the performance; the number of omissions which signifies the places where the red figures would land on the track which the subject did not observe; and the number of incorrect responses which means an apprehension of danger where none existed.

Another type of industrial test is for the selection of telephone operators. Both methods referred to have been employed for this purpose. McComas employed the method of a miniature model, constructing a miniature switchboard which enabled the experimenter to put the candidates through actual test calls and responses, during which performance speed and accuracy were measured. McComas believed that accuracy of aim or motor co-ordination was essential to the successful manipulation of a switchboard, and for the purpose of testing that factor he adopted a test, called the dot-striking test, a test which was originally devised by McDougall. In this test a sheet of white paper is stretched across a kymograph drum, and on the paper are eight rows of 120 red dots, each 1.5 mm. in diameter, and the dots arranged in an irregular fashion. The drum is revolved and as it does so the dots are visible through a horizontal slit. The subject is asked to strike each dot with a blunt soft pencil as the paper revolves, the speed of revolution being such that the subject can only succeed by putting forth his maximum effort. Whipple's criticism of the test is that in the subjects on whom he tried it there was a decided tendency for them to lapse into automatic performance. Other investigators have used the test in a modified form, among whom is McComas whose purpose was, as already indicated, to measure the accuracy of motor co-ordination which would be required for success in telephone operators.

Münsterberg also devised a test for the selection of telephone operators. He made a careful analysis of the psychological

processes involved in successful performance. The work is immensely taxing on the endurance and attention of the operator in a busy exchange. Most of the companies have found that the average operator cannot handle more than 225 calls per hour, though occasionally an operator is found who is able to answer more than 300 calls in an hour. In short periods operators have even attained the rapidity of 10 calls in a minute. Where the business of an exchange is very great it means that the element of fatigue has to be reckoned with, and that hygienic conditions must also be cared for. The inability of keeping the human nervous system at such a high point of tension for prolonged periods has to be recognized, if confusion is to be avoided and the health of the operators attended to. At the same time the psychologist who is engaged in testing candidates who are likely to succeed at this operation must bear in mind the concentration of attention at high pressure which is demanded, the fatigue which is likely to set in, and the accuracy demanded to avoid confusion.

Münsterberg was requested by a telephone company to study the mental requirements of employees, and began with an intensive study of thirty candidates. First he examined them with reference to psychophysical functions, including length of the fingers, rapidity of breathing, rapidity of the pulse, acuity of vision, acuity of hearing, distinctness of pronunciation, memory span, power of attention, general intelligence, accuracy and rapidity of responses. Psychological group tests were tried after which he turned to individual tests. The card-sorting test was given. Another test was given similar to the dot-striking test, small crosses being substituted for the dots, and the subject being asked to strike the crossing point with his pencil. This test was to measure ability such as is demanded in hitting the right holes in the switchboard in the telephone office. Another test was one in the cancellation of letters from the page of a newspaper in the belief that this operation involves an ability which functions also at the switchboard, though there directed to different material, namely concentrated attention. It will be seen that this investigator thus utilized several tests, and not one specialized test for selection for telephone service.

Many other attempts have been made at vocational tests. A form of the substitution test, which has been described in one form—the digit-symbol test—in the chapter on Performance Tests, has been utilized by some experimenters. Speed of improvement is the important element which is observed, and this is taken as indicative of the processes involved in business correspondence, stenography and type-writing. It has been ascertained that there is a fair degree of positive correlation between performance in these occupations and the substitution test. Other tests[1] have been

[1] See Hollingworth: Vocational Psychology, pp. 112—114.

used to measure ability in type-writing, and positive correlations obtained between actual performance and tests for memory span, tactual sensibility, muscular sensibility, sustained attention, and equality of strength in the two hands.

I need only refer to some of the other vocations upon which experimental work has been done in the way of devising tests as measurements of ability in performance. It will indicate the possibilities connected with vocational psychology, a science as yet in its infancy. The vocations include salesmanship, signalling, factory labour, music, clerical work, as well as those to which references have been made.

Another direction in which vocational psychology has moved is in administering tests of a wider range in order to discover vocational fitness without the specific purpose in view of making selections for a particular occupation, or at other times even with some specific vocation in view. Professor C. E. Seashore had made suggestions, e.g., toward a vocational psychograph with special reference to ability in singing. This psychograph[1] is a record of the measurements of the following abilities :—

I. Sensory—
  A.—Pitch—
    (1) Discrimination.
    (2) Survey of register of discrimination.
    (3) Tonal range.
    (4) Timbre discrimination.
    (5) Consonance and dissonance.
  B.—Intensity—
    (1) Sensibility.
    (2) Discrimination.
  C.—Time discrimination for short intervals.

II. Motor—
  A.—Pitch—
    (1) Striking a note.
    (2) Varying a tone.
    (3) Singing intervals.
    (4) Sustaining a tone.
    (5) Registers.
    (6) Timbre
      (*a*) purity,
      (*b*) richness,
      (*c*) mellowness,
      (*d*) clearness,
      (*e*) flexibility.
    (7) Plasticity ; curves of learning.

[1] See Hollingworth : Vocational Psychology, pp. 93—96, 294, 295.

B.—Intensity—
(1) Natural strength and volume of the voice.
(2) Voluntary control.

C.—Time—
(1) Motor ability.
(2) Transition and attack.
(3) Singing in time.
(4) Singing in rhyme.

III. Associational—

A.—Imagery—
(1) Type.
(2) Rôle of auditory and motor imagery.

B.—Memory—
(1) Memory span.
(2) Retention.
(3) Redintegration.

C.—Ideation—
(1) Association type and musical content.
(2) Musical grasp.
(3) Creative imagination.
(4) Plasticity; curves of learning.

IV. Affective—

A.—Likes and Dislikes — character of musical appeal.
(1) Pitch, timbre and harmony.
(2) Intensity and volume.
(3) Time and rhythm.

B.—Reaction to Musical Effect.

C.—Power of Interpretation in Singing.

V. Supplementary Data—biographical information, musical training, temperament and attitude, spontaneous tendencies in pursuit of music, general education and non-musical accomplishments, social circumstances, and physique.

Attempts have been made by some investigators to make enumerations of the characteristic abilities and motives and interests which are required for different occupations. Schneider has enumerated the following points concerning which observations should be made in a study designed to determine a subject's vocational fitness:—

(1) Physical strength; physical weakness.
(2) Mental; manual.
(3) Settled; roving.
(4) Indoor; outdoor.
(5) Directive; dependent.
(6) Original (creative); imitative.
(7) Small scope; large scope.

(8) Adaptable; self-centered.
(9) Deliberate; impulsive.
(10) Musical sense.
(11) Colour sense.
(12) Manual accuracy; manual inaccuracy.
(13) Mental accuracy (logic); mental inaccuracy.
(14) Concentration (mental focus); diffusion.
(15) Rapid mental co-ordination; slow mental co-ordination.
(16) Dynamic; static.

Hollingworth's criticism of this enumeration is that "the paired adjectives probably afford truer descriptions of various types of work than they do of types of individuals." [1]

Another enumeration was made by Münsterberg with reference to four specific vocations, the enumeration including abilities required, personal motives, and social interests.

Occupation.	Domestic worker.	Architect.	Physician.	Journalist.
Abilities required.	Joyful work ...	Aesthetic sense.	Social dealing ..	Sociability
	Energy ...	Imagination ...	Energy ..	Energy
	Patience ...	Industry ...	Discretion ...	Memory.
	Teaching ...	Drawing ..	Tact	Accuracy
	Economy ..	Modelling ...	Judgement ...	Judgement
	Physique ...	Specification	.....	Observation.
	.. ..	Employment of men.	.. ...	...
	Housekeeping.	Architecture ...	Dissection ..	Typewriting.
	Sewing ...	Engineering ...	Microscopical observation.	Quick expression.
	Cooking ...	Heating ...	.....	..
	Nursing ..	Ventilating ...	Psychotherapy.	Forceful style.
	House furnishing.	Construction ...	Clinical activity.	.....
	... ..	.....	Surgical technique.	.. ...
Implied personal motives and social interests.	Morality ...	Honour ...	Honour ...	Honour.
	Beauty ..	Beauty ...	Truth ...	Truth
	Position ...	Position ...	Position ...	Influence.
	Support ...	Fees ..	Fees	Salary.
	Home life ...	Comfort ...	Influence ...	Progress.
	Family welfare.	Progress ...	.. ..	.....
	Comfort of community.	Housing ...	Welfare of community.	Politics.
	...	.....	Health	Education.
	Family comfort.	.....	Prevention of disease.	Information.
	... ..	.....	.....	Entertainment.

B.—Measures of Character.

There are some obvious limitations to the intelligence tests. Among them is that they do not measure the emotions. There are also certain moral or social qualities of personality which cannot be measured by any means that have been devised as yet. But

[1] Op. cit., p. 106.

there is more correlation between intelligence and traits of character than some might imagine at first. The fact is that character and personality are both of them terms of great complexity, even as intelligence is. All three words are used by various persons with wide divergences of inclusiveness or exclusiveness. For example Thorndike[1] in an article in *Harper's Magazine* on *Intelligence and Its Uses* makes the term a very inclusive one. He describes intelligence as of three kinds: the abstract intelligence, the mechanical intelligence, and the social intelligence. When he speaks of social intelligence he means to include practically all of the so-called character traits. Another author, Fernald [2] suggests likewise that intelligence is variable, variations being not only quantitative, i.e., of degree, but also qualitative by which he means the character traits.

If the above interpretation of intelligence be valid, then we may expect tests of intelligence to give some indication of character also. And practically all of the investigators claim that the tests have been a help in that direction. In other words they have discovered a positive correlation between traits of character and intelligence. Terman, e.g., made a study of the extent to which intellectually gifted pupils possess the following personal and moral traits and found that there was a positive correlation in every case: sense of humour, power to give sustained attention, persistence, initiative, accuracy, will power, conscientiousness, social adaptability, leadership, personal appearance, cheerfulness, co-operation, physical self-control, industry, courage, dependability, self-expression through speech, intellectual modesty, obedience, popularity among fellows, evenness of temper, emotional self-control, unselfishness, and speed. Terman [3] found in the case of the sense of humour a correlation with intelligence in the case of gifted children of .58; and in the case of speed, the last in the list a correlation of .28. This author claims that he can roughly predict the intelligence quotient from the average of these 24 traits.

Professor A. T. Poffenberger of Columbia University, in an article in the *Journal of Philosophy* expresses the faith there are greater possibilities in this direction than anything so far accomplished. He says:

"With some modification of content, method of administration, and with supplementary scoring such a test as the Army Alpha might be made to yield measures of neatness, accuracy, speed of decision, freedom from inertia, assurance, willingness to take a chance, tenacity or perseverance, honesty, etc. The total score from such a test would give a measure of efficiency or competence.

---

[1] 1920, Vol CXL, pp 227—235.

[2] *Journal of Abnormal Psychology*, 1920, Vol. XV, pp. 4 ff.

[3] See Terman: The Intelligence of School Children, p. 58.

By proper weighing of the different ingredients of the total score, measures could be provided for different occupations . . . Such a combined measure of intelligence and character, if used for vocational purposes, would prevent the waste of high grades of intelligence in positions where it is not needed and would enable those of low intelligence to be located where their capacity would be adequate and where their character traits would make them successful. . . . To refuse an occupation in business and industry to all persons with an intelligence under seventy per cent of normal, without examination of their character qualities, may sometime appear to be one of the greatest of human and economic wastes."[1]

The truth is that, important as intelligence is, it is not the sole requisite for useful citizenship. We noted in the case of the United States Army that there was plenty of work for men of inferior intelligence and even for a great many of the men of very inferior intelligence. Out of the million and three quarters men who were examined, only 7,800 or one-half of 1 per cent were recommended for discharge, whereas nearly 20,000 men were useful, though of very inferior intelligence. Otis made an investigation as to the correlation between success as a mill worker and intelligence and found it nil.[2] His conclusion was that intelligence was not a requirement of a worker in a modern silk mill, and hazards the possibility that the qualities needed may be stolidity, patience, inertia of attention, regularity of habits, etc.

Fernald has dealt with the same question in the article in the *Journal of Applied Psychology* to which reference has been made. He describes the cases of two young men. The one was an employer's confidential clerk with a creditably high intelligence quotient, but whose fast living occasioned failures leading to his forging his employer's signature three times. The other was a farm boy who scored only 39 of an I. Q., but who did his work faithfully and behaved himself well. "The findings of intelligence tests only in these two cases are that *A* is of at least ordinary intelligence while B is an imbecile. The findings of character study only are that *A* is legally an offender, an economic parasite and a social menace, while *B* is law abiding, a producer and no menace. Consideration of both fields of inquiry affords a far broader and more illuminating and therefore truer basis of comparison than is available from the consideration of either field alone. In fact, conclusions drawn from investigations in either field to the exclusion of the other is misleading."

Fortunately in the majority of instances we have a positive correlation between tests of intelligence and judgements of

[1] Measures of Intelligence and Character, in Vol. XIX, No. 10, May 11, 1922, pp. 261—266.

[2] See *Journal of Applied Psychology*, 1920, Vol. IV, pp. 339—341.

character. So that we do not have the contrast which Fernald finds in these concrete instances repeated very often in actual experience. But the fact that there are even a few people of high mentality and low character and another few with low mentality and good character means that an injustice would be done in both cases, if we were to determine the places into which they should be put vocationally purely on the basis of intelligence tests.

Professor J. McK. Cattell, in his *Home Scientificus Americanus*, has attempted an inventory of character traits, as follows:—

Physical health.	Energy.	Unselfishness.
Mental balance.	Judgement.	Kindliness.
Intellect.	Originality.	Cheerfulness.
Emotions.	Perseverance.	Refinement.
Will.	Reasonableness.	Integrity.
Quickness.	Clearness.	Courage.
Intensity.	Independence.	Efficiency.
Breadth.	Co-operativeness.	Leadership.

Dr. F. L. Wells has made a study of this problem on the basis of the work begun by Cattell and others. He has made an inventory[1] of fourteen phases or aspects of human personality, and in connection with each phase has suggested certain questions, clues and features by which their presence or absence may be diagnosed. Under these fourteen main traits he has in all about ninety-five sub-traits. The following is his outline:—

1. Intellectual processes (5 sub-topics).
2. Output of energy (4 sub-topics).
3. Self-assertion (7 sub-topics).
4. Adaptability (5 sub-topics).
5. General habits of work (5 sub-topics).
6. Moral sphere (6 sub-topics).
7. Recreative activities (16 sub-topics).
8. General cast of mood (3 sub-topics).
9. Attitude towards self (4 sub-topics).
10. Attitude towards others (7 sub-topics).
11. Reactions to attitude towards self and others (12 sub-topics).
12. Position towards reality (5 sub-topics).
13. Sexual sphere (9 sub-topics).
14. Balancing factors (6 sub-topics).

The analyses of Cattell and Wells show that there is a great deal of difficulty involved in the measurement of character. First of all, character is so complex that the task of analysis is itself enormous. Furthermore there is so much inter-penetration between the qualities which make up the complex that it is difficult to discover what predominates in some instances. In addition there is then the task of devising tests which shall be indicative of these,

[1] See *Psychological Review*, July 1914. *The Systematic Observation of Personality.*

qualities. The indications seem to point to only limited possibilities in this direction, because character is a complex of moral qualities and attitudes, elements that do not readily yield to the mechanistic processes. Yet even if it may not be possible to attain to any great success in measuring the amount or degree in which these qualities are present in an individual, there seems no ground for supposing that they may not be discovered to be present or absent. And, if Thorndike's theses that "whatever exists at all, exists in some amount," and "anything that exists in amount can be measured" be true, perhaps we may hope for the day to come when we shall be able to measure traits of character by quantitative standards.

Dr. June E. Downey of the University of Wyoming has devised a test which is designed to afford an index to certain character traits. The test is called the "Downey Individual Will-Temperament Tests," and is a step in the direction of the measurement of character, though not as satisfactory as psychologists hope to achieve in the future. It must be admitted that Professor Downey has devised tests which are well adapted to indicate the presence or absence of certain traits of temperament and will, though it may be questioned as to the accuracy of the measuring devices which are used. There are thirteen tests in the series. The first one presents to the examinee a list of paired words which express traits of temperament in contrast, such as "careful-careless," "industrious-lazy," "vain-modest," "hasty-deliberate," and "extravagant-thrifty," and the examinee is asked to grade himself on each trait by checking only one of each pair. The subject has the privilege of qualifying, if he desires, by the use of percentages. The examiner does not give the test for the sake of securing the subject's own estimate of himself, but to determine the speed with which the person makes decisions in general. So that the significant things are the time required and the reasons for any delay. The second test is one in which the subject is required to sign his name as rapidly as possible. By a comparison with his normal rate it is possible to detect tendencies to procrastinate or adopt an unnecessarily slow pace when not under pressure. In the third test the subject is requested to write his name as slowly as possible, the purpose being to discover what ease and success the subject possesses for modification and adjustment. Dr. Downey says that "a very high score probably indicates some finesse in the handling of personal relations, or dramatic ability."[1] The fourth test consists of showing the person two envelopes in which he is told there are different mental tests one of which is easy and the other difficult, and is asked to choose one of them without being informed which envelope contains the easy and which the hard one. Nothing is done with this at the time except that the examiner without the

[1] Manual of Directions, p. 20.

knowledge of the examinee records the choice. But later, as test XI, the examiner returns to this by picking up the envelopes and asking the subject which he has chosen after which he contradicts him, the object being to determine by his reaction to contradiction what degree of assurance he has, and to what extent he is willing to accept the responsibility for his decisions. Test V is one of co-ordination of impulses, the examinee being required to write the words "United States of America" as rapidly as possible at the same time writing within a small space. The author considers that this test is a measure of one's ability to handle a complex situation, such as may be required in driving an automobile quickly and carefully through a crowded street. The test is calculated to indicate the person's ability to make inhibitions and to avoid explosive actions. Other writing tests are utilized to bring out certain temperamental traits. In tests six to nine inclusive the phrase "United States of America" has to be copied (i) at the usual style and speed, (ii) as rapidly as possible, (iii) as slowly as possible, (iv) in a disguised hand, (v) as exactly as possible to two models. In the tenth and twelfth tests the person is required to write his own name (i) eyes closed, usual style and speed, (ii) while counting rapidly by 3's, eyes open, (iii) while counting rapidly by 3's, eyes closed, (iv) beginning at the 7th tap of a pencil, eyes closed, counting rapidly by 2's, and (v) at usual speed, eyes closed. These exercises are all designed to discover certain temperamental traits such as motor inhibition or the patience required in the face of a disagreeable piece of work (by writing exceedingly slow), the ability to persevere in situations that require a departure from the routine way of acting (as in disguised handwriting), one's interest in exacting details which is requisite for success in so many vocations (as in copying the presented models), and the amount of energy and the person's ability to carry out instructions in spite of distractions (as in writing while counting). In one of the writing tests an effort is made to measure the subject's ability to resist opposition by compelling him to write while an obstacle is placed in front of his pen. Success in this test is an indication, according to the author, of a "man with fighting qualities." "The unagressive person evades the issue or gives up".

Dr. Downey obtained norms for her test on the percentile basis. That is, she gave a score value of from one to ten for each test, and arranged the scores of her subjects so that ten per cent of the persons tested obtained each score. On this basis she constructed what she called "the will-profile" of each testee by the graph method. This will-profile ought to enable a person to tell at a glance the dominant traits of temperament in any person who has been tested. But the numbers so far tested are rather small for one to be guided, except in a general way, by them. The main thing is that a beginning has been made which points the way to a real possibility in measuring character traits.

CHAPTER VIII.

TESTS OF ACHIEVEMENT.

We are concerned in this subject with the bearing of mental tests on the classification of pupils and the organization of a school. The question of classifying pupils for school work is one in which the teacher is most vitally interested. There are three ways in which it has been done. First, children may be classified on the basis of intelligence. There are some ardent advocates of intelligence measurements who claim for them that they are a sufficient criterion without anything else on which to organize a school. There are others who equally oppose the intelligence test as a basis for classification. But there is a safer middle ground to take.

The intimate connection between mental age and school performance cannot now be questioned. When a pupil fails to make the progress that he should, the first thing to do is to administer a test of intelligence to ascertain the quality of work of which he is capable. Terman, Whipple, McCall, Dickson and others have shown conclusively that there is a close correlation between mental age and the quality of school work. There is more validity in the intelligence tests than there is in the judgement of the teacher, and they ought to be given to all pupils as an aid to the teacher in classification. Terman has collected statistics to show that in practically every grade where intelligence tests have not been used one can find 25 per cent of the pupils who ought to be in a lower grade and an equal number who ought to be in a higher grade, and that in almost every grade there are pupils ranging in mentality from eight to fourteen. These are irregularities which can readily be corrected by a proper observance of intelligence tests.

In the second place, children are classified on the basis of the marks which are given by teachers — a pedagogical basis. Obviously the teacher's judgement is not of any use when a child first enters school or when he enters from another school. Terman, Whipple, and McCall have brought forth much evidence to show that the judgement is inaccurate even when he knows his pupils well, because he frequently fails to take into account some of the factors, such as the relationship of chronological age to grade. The marks of a teacher in ordinary class examinations are of value, especially when they are the only records available, but they are also subject to the error of lack of standardization.

The third basis for classification is the educational test which is a standardized test of achievement. McCall lays it down that when educational tests are to be used as a basis for the classification of a school three points should be observed. These are :

(i) The test should be uniform for all grades being classified or reclassified. That means that tests must be used which are

capable of being used in all grades regardless of their being lower or higher. Otherwise it is not possible to make legitimate comparisons.

(ii) The test ought to yield a single score. A double basis for scoring such as for speed and accuracy is difficult, especially for those who are inexperienced in administering tests.

(iii) The test should be designed so as to measure an important phase of the work of the school. As a rule, different tests should measure attainments in different subjects.

The fundamental principles which must be remembered in classifying pupils are two : (i) pupils of equal status ought to be placed in the same class ; and (ii) pupils should be put together who are likely to make progress at an equal rate. It is simply the application of a principle of logic to say that homogeneity should be a characteristic of a class. And the types of homogeneity that are most significant in classifying pupils in a school are the two that have been indicated, viz., educational status and ability to make progress. There are many times when both of these matters are shamefully neglected and when other inadequate bases are substituted. Professor Judd has given a summary of such ill-advised influences in his *Introduction to the Scientific Study of Education* :

" Sometimes the school allows a pupil to move up a grade or class, although it is known that he has not done the work below, because the parents of the child have influence and it does not seem safe to antagonize them.

" Sometimes the pressure of numbers in the lower grades or classes is so great that the teacher sends a pupil on in order to make room for the younger pupils, even when it is evident that the pupil will not be able to carry the higher work.

" Sometimes the teacher in a given grade is anxious to unload the backward or disorderly and therefore incompetent pupil on someone else, and since the only open road is into the next higher grade, the child is sent on.

" Promotion is sometimes controlled by the calendar. Because the date for closing the schools has arrived, and the long vacation is at hand, pupils are declared to have completed the work, whether they have or not.

" Sometimes it is more or less explicitly argued that the backwared pupil is larger than the other children of like intellectual attainments and he should therefore be sent to the upper-grade room where the seats are larger."[1]

Besides affording a basis for the classification of pupils the standardized tests serve a second useful purpose, viz., in diagnosis both of the ability and of the peculiar difficulties of a pupil. There are two types of diagnosis, viz., the general diagnosis which is

[1] Pp. 109 and 110.

concerned with analyzing the subject's initial condition, and the detailed diagnosis which is a more careful analysis of his specific abilities and defects. The purpose of the diagnosis is to serve as a guide for futher instruction with a view to correcting defects. Individual treatment is needed, of course, for the purposes of diagnosis and correction. It will often happen that a child with quite marked ability fails in some particular operation, such as an arithmetical operation, because he has a defective understanding of the nature of the process, due perhaps to a gloss in the teaching or may be to divided attention when the subject was taught. One of the most valuable uses of the standardized test is that it may be used as an instrument for diagnozing pupil's particular difficulties and at the same time revealing defective instruction.

Professor McCall has a very fine discussion[1] of the various methods which may be employed for diagnostic purposes. The list is as follows:—

(i) Introspection by the pupil.—Pupils very frequently know the exact location of their difficulties, and sometimes the causes as well.

(ii) Observation of normal work.—This is a method which a teacher employs regularly, and frequently gives the key to the situation. As one's experience in teaching grows, his ability in diagnosis by observation should keep pace.

(iii) Oral tracing of process.—There are difficulties which only come to light with a series of questions as to the process involved so as to reveal where the difficulty is.

(iv) Analysis of test results.—Many of the tests of attainment have been especially designed with a view to enabling the instructor to locate the difficulty.

(v) Developmental history.—Many difficulties which a pupil experiences have existed for a considerable period, so that the history of the pupil's development is as necessary to the psychologist as the history of a patient is to a physician.

(vi) Contrast of opposites.—It sometimes happens that a teacher is able to diagnoze a pupil's difficulties by contrasting him with another one who succeeds in the same operation.

(vii) Complete analysis of ability.—"A complete and thorough analysis of the sensory, mental, and motor processes involved in a given ability is the last resort of the diagnostician." This method too means a thorough use of tests as technique.

In the question of diagnosis a good deal of valuable work has been done by various educational psychologists and the results are summarized in various places. Professor S. A. Courtis in his *Teacher's Manual for the Standard Practice Tests* has treated the matter of arithmetical defects, pointing out causes and suggesting ways of remedying them. They may be summarized as follows:—

---

[1] How to Measure in Education, pp. 89—102.

(i) Movements slow and deliberate but steady.—This may be due to bad habits or to retarded neural action. There is no remedy equal to practice for this defect.

(ii) Movements rapid but variable, indicating nervous strain.—The cause may be sought and remedied in the environment or conditions of work.

(iii) Progress irregular.—This may be due to lack of controlled attention or to lack of knowledge of the conditions. A teacher must realize that "inattention" is a psychological anomaly, that the real trouble is attention being diverted, and should seek to establish conditions which will prevent the division of attention.

(iv) Pupil stopping to count by the fingers or dots on paper or other mechanical aids.—The only remedy is a proper learning of the combinations.

(v) Adding each first column correctly but frequently missing on the second or third columns.—This is due to weak memory habits in carrying and may be corrected by attention to that process.

(vi) The time required for working problems increases either steadily or irregularly.—An indication of the fatigue factor which is very hard to remedy and needs special attention to each individual case.

(vii) Habits apparently good and work steady, but the child answers incorrectly.—Requires a careful study of the process step by step with a view to discovering the place where the child goes wrong.

A careful diagnosis involves a study of the mental operations involved in any process. Professor Leta S. Hollingworth has made such an analysis [1] for the operation of spelling, based on the results of experiments in the Teachers' College at Columbia University. The processes involved in poor spelling include the following:

(i) Sensory processes—defective hearing or defective vision is likely to result in errors in spelling.

(ii) General intelligence—general intellectual weakness may be the cause of poor spelling.

(iii) Faulty pronunciation—this may be due to faulty auditory perception, or to the inability to articulate properly.

(iv) Wrong associations due to faulty visual perceptions.

(v) Failure to remember or to retain impressions due to a short memory span.

(vi) The rational element, by which is meant an understanding of the meaning of a word.

(vii) Motor awkwardness and inco-ordination indicating a weak or slow response system.

---

[1] See Hollingworth and Winford: *The Psychology of Special Disability in Spelling*. Teachers' College Contributions to Education, No. 88, 1918.

(viii) Lapses due to carelessness and to weakness in concentration.

(ix) Transfer of habits previously acquired—a frequent cause of poor spelling, where a person begins to use a new language.

(x) Individual idiosyncraries—no general explanation.

(xi) Temperamental traits in which emotional factors play a greater part than intellectual.

A third purpose is served by the standardized test, namely the measurement of the results of teaching. As already indicated the discovery of a weakness in a child's operations may be due to either of two causes, a defective comprehension of the process for which the pupil is responsible, or one for which poor instruction is responsible. Here as elsewhere "the proof of the pudding is in the eating thereof." And the teacher will find the standardized test a most invaluable mechanism wherewith to check the efficiency of his own work. It must always be remembered that the pupil is the center of interest in education. The whole mechanism of education exists for no other purpose than to help him to make progress, and the worth of any detail in the system is measurable in terms of its usefulness in aiding the development of the pupil. On that basis the only criterion on which to judge of the worthfulness of a teacher is with reference to the pupil. The teacher who is able to influence the pupil in the direction of the normal unfolding of personality and his best progress is successful, and the one who fails in that fails in his vocation. The problem is how to select teachers for appointment and promotion, i.e., how to measure teaching. Certainly physical appearance, vivacity, attractiveness of personality, or even general intelligence are not the measures of good teaching. The standardized measurement which indicates the amount of progress which pupils have made under the direction of a teacher is the best criterion of success or failure in instruction.

When we use the phrase "standardized measurement" with reference to teaching we must take into account a number of factors. The time factor is one. In the measurement of progress it is only fair that pupils should be equated with reference to the length of time involved. The pupil factor is another. This is where the importance of the Intelligence Quotient comes in. Pupils of superior intelligence are capable of making progress at a more rapid rate than pupils of low intellectuality. Standardizing the test itself is a third factor, and that involves its application to a sufficiently large number of subjects to secure a median for a given grade or age, and a thorough testing of the test as a measure of ability or achievement. The estimation of a teacher's efficiency in instruction can be done fairly only under such well standardized conditions, and the tests are being used increasingly for such purposes.

We are familiar with the term, "Intelligence Quotient." Since the extension of the psychological tests to the realm of educational attainments a new term has been introduced which refers to the status of the educand in educational accomplishments. This term is "Educational Quotient." McCall has described[1] the method of computing the Educational Quotient. In measuring a school for purposes of classification it is necessary to administer a number of tests in the various subjects of instruction. These tests have to be administered according to standardized procedure as already described. The tests must then be scored and the computation of the individuals' scores obtained. Then the scores must be tabulated and the median computed. By the median is meant the score of the pupil whose score is such that there are fifty per cent who score higher and an equal number lower than he does. The next step is the tabulation of the norms for the tests and grades. Next comes the computation of the composite score for each subject, but that may be different from merely totalling his scores in the various tests because the tests may not be weighted proportionately. Hence the necessity of readjusting the composite score by making the weighted score proportionate to the other tests. It is necessary to take into account the average chronological age of the pupils in various grades. This has been done approximately, and it is known, e.g., that the average at which American children enter school is 80 months. Having determined the average number of months which pupils spend in a single grade it is now possible to determine the average chronological age for each grade. From the composite norm and the average chronological age it is possible quite readily to compute the educational age of any child. Thirteen months is the average which has been computed as the time spent by pupils in a grade. So that the norm composite in relationship to the educational age gives a person's educational age. Supposing a person scores 188 as a composite score, though only in the sixth grade. The table shows us that 188 is the norm composite for seventh grade pupils whose average chronological age is 167 months. We are able at once to fix the subject's educational age as 167 months. But we ascertain that the child's chronological age is 150 months. The Educational Quotient is computed by dividing the educational age by the chronological age. So this particular subject's E.Q. would be $\frac{167}{150} = 111$.

It is McCall's mature judgement that the E.Q. gives a more valuable criterion for school organization, if it has been calculated on the basis of a proper scale of educational tests, than does the I.Q. It is superior in the first place because it affords a basis for educational classification which must be the basis used in school organization. Its superiority also comes out in that it prevents

[1] How to Measure in Education, pp. 25—45.

pupils from skipping important parts of the school curriculum. It also prevents the skipping of certain parts of school work which are important for the unfolding of special abilities. If there is a wide disparity between mental and educational ages this can be remedied by appropriate instruction which will enable the child to advance educationally until he reaches the class which represents his mental level.

We shall now pass on to a perusal of some of the standardized tests which are in use. In the case of such tests the group method is the one employed, as it enables the examiner to examine a whole class at one time. Not only so, but the efficiency of the method depends on being able to construct norms for classes with which the individual may be compared, so that group testing offers an opportunity for collecting a large amount of data in a short time. The usual method is to have the test printed or cyclostyled with blanks in which the subject can record his answers, and also with score values indicated so that the examiner can speedily and accurately compute the score. In other cases cards are printed with the correct answers indicated and holes cut out which will fit exactly over the answers of the pupils, so that the examiner may place this correct answer key over the pupils' card and at a glance compare the pupil's performance with the correct one. This facilitates the scoring. Every fresh group of results that is obtained affects the average and the median, so that medians are being constantly adjusted and modified. Differences in educational systems in various countries makes it difficult to obtain standards that are internationally valid, and in some respects it may be that every country will have to work out its own, but in so far as the standards and medians can be made world-wide, to that extent we shall be able to increase the value of our educational comparisons.

I.—The Measurement of Arithmetical Abilities.

The popular notion is that a person is good or bad in arithmetic, but it has not occurred to most people that there is no one simple process involved in arithmetical operations. It is quite possible that a person may be good in one or more processes and not so in another or others. Whereas on the whole there is a very fair degree of correlation between abilities in the various operations, it does not follow that such is invariably the case. The fact is that arithmetical operations call for the function of many processes, each of which has its characteristic difficulties. A number of years ago Stone made an investigation of the matter[1], and concluded that arithmetical abilities were specific. So that the teaching of

[1] Stone, C. W.: Arithmetical Abilities and Some Factors Determining Them, 1908.

arithmetic involves the engendering of a number of specific abilities relatively distinct rather than a single ability, the word 'ability' signifying the rate and accuracy with which a subject performs a certain operation. On that basis it has been concluded that there are as many abilities as there are types of operations.

Psychologically speaking the functions are also complex The learning process which operates in such a case as the learning of the multiplication tables is one which involves several factors including visual memory with many subjects, association, attention keenness of observation, and facility in habit formation. In the addition of a column of figures there are several functions operative, but perhaps the most significant is the process of attention. The ability to add correctly long columns of figures depends for one thing upon the span of one's attention. This is itself a very complex matter as anyone knows who has studied the subject in psychology. It includes the question of interest and selection, the ability to discriminate and to make combinations, and the interweaving of the factors of shifting and sustaining attention. Attention is sometimes measured in the laboratory by experimental methods, but the addition of columns of figures of graduated length is as good a way of any of testing the span. It will be found that children will develop in their spans with observation and practice. In other words, the span of attention is educable. A person may increase his span in addition operations by constant practice, as a person may increase his facility in observation. Attention to attention will increase its power. One may readily discover the span of his own attention by observing the point at which fatigue sets in, as he adds a column of figures. In such a process it is necessary for one to hold in mind the partial sum until he has added the next figure. Frequently one will observe that there is a tendency to stop, a tendency to uncertainty sets in at about the same point in each column, and so he begins again. The point where such uncertainty sets in marks the fact that he has exceeded the span of attention.

An analysis of the operations with integers has been made by S. A. Courtis in his *Teacher's Manual for Courtis Standard Practice Tests* (1916). The following typical operations are differentiated:

Addition: (i) simple addition combinations such as 2 + 3;
(ii) single-column addition of three figures such as 4 + 3 + 7;
(iii) "bridging the tens," as 38 + 7;
(iv) column addition, seven figures;
(v) addition with carrying;
(vi) column addition with increased span, thirteen figures to the column;
(vii) addition of numbers of different lengths;

Subtraction: (i) simple subtraction combinations, such as $4-3$;
(ii) subtraction of 9 or less from a number of two digits, without "borrowing";
(iii) same as the second, but with "borrowing";
(iv) subtraction of numbers of two or more digits involving borrowing;

Multiplication: (i) simple multiplication combinations, such as 5×4;
(ii) multiplicand two digits, multiplier one digit, and no carrying, such as 34×2;
(iii) same as number two, but with carrying;
(iv) long multiplication, without carrying, such as 23×41;
(v to viii) zero difficulties, four types, e.g.,

560	807	617	703
40	59	508	60

(ix) long multiplication, with carrying.

Division: (i) simple division combinations, such as $4 \div 2$;
(ii) simple division, no carrying, such as $36 \div 3$;
(iii) same as number (ii), but with carrying;
(iv) long division, no carrying;
(v and vi) zero difficulties, two cases, e.g.

$$\frac{48990}{71} = 690 \qquad \frac{9362}{31} = 302;$$

(vii) long division, with carrying, "first case, where the first figure of the divisor is the trial divisor, and the trial quotient is the true quotient," e.g.,

$$\frac{4536}{63} = 72;$$

(viii) "second case, where the trial divisor is one larger than the first figure of the divisor, and trial quotient is the true quotient," e.g.,

$$\frac{3087}{49} = 63;$$

(ix) "third case, where the first figure of the divisor is the trial divisor, but the true quotient is one smaller than the trial quotient," e.g.,

$$\frac{5607}{63} = 89;$$

(x) "fourth case, where the first figure of the divisor must be increased by one to obtain a trial divisor, and the second trial quotient

must be increased by one to get the true quotient" e.g.,

$$\frac{2844}{36} = 79$$

In commenting on these findings of Courtis, Professor W. S. Monroe says: "Each of these types of examples requries a specific habit or automatism. To be sure, certain elements, such as the fundamental combinations, are common, but careful analysis will show that ability to do examples of one type is different from that required to do another. Not only will a careful analysis reveal this fact, but it has been repeatedly demonstrated by carefully conducted investigations. In addition to the specific automatisms which are required for the four fundamental operations with integers, a number of other automatisms are required for operations with fractions both common and decimal. At present we have only a partial analysis of the example in these fields, and for that reason it is not possible to state what types of examples are within the range of school work.

"The significant characteristics of these abilities or automatic responses are the rate or speed of performance, the accuracy of performance, and the accuracy of the response. Thus, the measurement of arithmetical abilities involves determining both at what rate a pupil is able to do examples of the elemental types, and how accurate his answers are. This is accomplished by having him do examples of a given type for a specified time. From his test paper his rate and per cent of examples correct may be determined. These two quantities represent the measure of his ability to do this type of example.

"Strictly speaking, the number of examples done and the per cent of examples correct is a measure of the pupil's performance rather than of his ability. A pupil's performance is affected by many factors such as his emotional status, physical condition, light, temperature, and the like. Or, it may be that a pupil does not try to do his best on a given test. A pupil's ability can only be inferred from his performance, but when conditions are properly controlled, such inference is reliable in all except a few cases. In order to avoid an awkward form of statement and because the practice is general, we shall speak of a score as a measure of a pupil's ability."[1]

There are several tests of arithmetical abilities which are now in use. The following may be mentioned as typical:—The Courtis Standard Research Tests, The Stone Reasoning Test, Monroe's Diagnostic Tests, Woody's Arithmetic Scales, The Cleveland Survey Arithmetic Tests, Kansas Diagnostic Tests in Arithmetic,

[1] Monroe, W. S.: *Measuring the Results of Teaching*, pp. 113, 114 and 114 n. Boston: Houghton Mifflin & Co., 1918.

Boston Research Tests in Fractions, Ballard's Tests in Arithmetical Reasoning, Burt's Tests in Mechanical Arithmetic, Starch's Arithmetical Scale, etc. In addition to the tests themselves there is a considerable amount of literature[1] already available dealing with the tests and with the abilities which the tests are designed to measure.

Arithmetical problems bring into play the reasoning processes. Reasoning has been defined by Woodworth as "mental exploration" as "distinguished from motor exploration of the trial and error variety."[2] It is an explorative process in which the subject attends to a definite problem, thinks it through instead of mechanically searching for a solution, and calls upon the experiences of the past for light on the present problem. It is logically a process of inference, because there is no presentation of objects to the senses. It is a mental manipulation of data in which a response is mentally determined on the basis of mental stimuli. Arithmetical problems are well calculated to test that type of ability, an ability which is educable and concerning the attainment of which the educator is interested. The higher up the scale in school work a child may be, the greater the necessity that the measurement of arithmetical ability should be so designed as to call into play reasoning. The Stone Reasoning Test is a test in which the subject is allowed fifteen minutes for the solution of twelve problems, and since it has been administered to a great many subjects it has been possible to standardize the performances according to grades. The following is the form of the Stone Reasoning Test in a form which may be more suitable to India. I have kept the problems the same, simply substituting Indian for American terminology and currency.

THE STONE REASONING TEST (ADAPTED).

(*Time—exactly* 15 *minutes*).

School— Grade— Name of pupil—

Solve as many of the following problems as you have time for; work them in order as numbered:—

Problem value.	Problems.
1·0	1. If you buy two writing pads at As. 7 each, and a book for Rs. 2-8-0, how much change should you receive from a Rs. 5 note?
1·0	2. Ramaswami sold 4 newspapers at As. 2½ each. He kept ½ of the money, and with the other ½ bought more papers at Anna 1 each. How many did he buy?

[1] See the bibliography at the end of Chapter IV. The Measurement of Arithmetic, in Wilson and Hook: *How to Measure*, New York: The Macmillan Co., 1921.

[2] Psychology. A Study of Mental Life, p. 462.

Problem value.	Problems.
1·0	3. If Krishnayya had 4 times as much money as Venkatayya, he would have Rs. 16. How much money has Venkatayya ?
1·0	4. How many pencils can you buy for Re. 1–8–0 at the rate of 2 for As. 3 ?
1·0	5. The uniforms for a football eleven cost Rs. 7–8–0 each, and the boots cost Rs. 6 per pair. What was the total cost of uniforms and shoes for the eleven ?
1·4	6. In the schools of a certain city there are 2,200 pupils ; $\frac{1}{2}$ are in the elementary grades, $\frac{1}{4}$ in the lower secondary grades, $\frac{1}{8}$ in the upper secondary grades, and the rest in the night school. How many pupils are there in the night school ?
1·2	7. If $3\frac{1}{2}$ tons of wood cost Rs. 21, what will $5\frac{1}{2}$ tons cost ?
1·6	8. A newsdealer bought some magazines for Rs. 3. He sold them for Rs. 3–12–0 gaining As. 3 on each magazine. How many magazines were there ?
2·0	9. A boy spent $\frac{1}{8}$ of his money for tram fare and three times as much for clothes. Half of what he had left was Rs. 2–8–0. How much money did he have at first ?
2·0	10. Two tailor's chokras receive Rs. 17–8–0 for sewing shirts. One makes 42 and the other 28. How shall they divide the money ?
2·0	11. A certain Chetti paid one-third of the cost of a building ; his partner received Rs. 500 more annual rent than the Chetti. How much did each receive ?
2·0	12. A goods train left Madras for Madura at 6 o'clock. The mail train left on the same track at 8 o'clock. It went at the rate of 40 miles per hour. At what time of day will it overtake the goods train if the goods train stops after it has gone 56 miles ?

The method of scoring is to give to each problem solved correctly the value indicated in the margin. Dr. Stone has issued the following table of norms which are based on the median scores obtained after using the test in many cities :—

Grades.	Standards.
5	Score of 5·5, reached or exceeded by 80 per cent, 75 per cent accuracy.
6	Score of 6·5, reached or exceeded by 80 per cent, 80 per cent accuracy.
7	Score of 7·5, reached or exceeded by 80 per cent, 85 per cent accuracy.
8	Score of 8·75, reached or exceeded by 80 per cent, 90 per cent accuracy.

The Courtis Arithmetic Tests (Series B) consists of tests in addition, subtraction, multiplication, and division, constructed in such a way that each problem is of equal difficulty to every other. Twenty-four problems in addition are given with a time limit of 8 minutes; 24 in subtraction for four minutes; 25 in multiplication for 6 minutes; and 24 in division for 8 minutes. The following are samples from each test:—

Test No. 1.—Addition.

927	297	136	486	384	176	277	837
379	925	340	765	477	783	445	882
756	473	988	524	881	697	682	959
837	983	386	140	266	200	594	603
924	315	353	812	679	366	481	118
110	661	904	466	241	851	778	781
854	794	547	355	796	535	849	756
965	177	192	834	850	323	157	222
344	124	439	567	733	229	953	525

Test No. 2.—Subtraction.

107795491	75088824	91500053	87939983	160620971
77197029	57406394	19901563	72207361	80361837

Test No. 3.—Multiplication.

8246	3597	5739	2648	9537	4258	7593
29	73	85	46	92	37	640

Test No. 4.—Division.

25)6775 94)85352 37)9590 86)80066 73)58765

The Courtis tests are so devised that an instructor will have no difficulty in administering them, even though he may have had no previous experience, if he but follows the instructions. Great care is taken about the time element, because speed as well as accuracy is taken to be necessary in measuring the results of teaching in arithmetic. It is emphasized that all must begin at the same time and all must stop at the same time. In beginning, the printed test papers are always arranged on the desks ready for work, while the subjects with pencils in hand maintain the attitude of asking a question with their hands raised. Then when the signal is given the hands are brought down and work begun simultaneously. When the signal to stop work is given they must cease, even if in the middle of writing a figure, and put their hands up again. The correct answers are read and the children are allowed to check the number correct and the number wrong, and write in their total score. By having the papers exchanged for scoring a good deal of time may be saved the instructor, whereas he may check up a certain amount afterwards to make sure that instructions were followed correctly and that the scoring was done properly.

The significance of the results can be realized only as they are compared with the standards which were designed by Courtis on the basis of the experiments which he carried through with the tests. Wilson and Hoke in *How to Measure* (pp. 58—74) and Monroe in his *Measuring the Results of Teaching* (pp. 119—131) give a number of statistical tables which deal with the results obtained both by Courtis himself and by other investigators who have made use of his tests. A record is made of the number of problems attempted as well as those done right. The results are arranged in accordance with a grade-scale. The following table will illustrate from Courtis 1916 investigations :—

Grade.	Addition.	Subtraction.	Multiplication.	Division.
III	4	5	0	0
IV	6	7	6	4
V	8	9	8	6
VI	10	11	9	8
VII	11	12	10	10
VIII	12	13	11	11

Standard of accuracy, 100 per cent.

Ballard's most serious criticism of the Courtis tests is that they are standardized on a grade-scale, whereas that makes them insular and prevents comparison with children in other countries where the grading is different. He feels that the best way to obviate that difficulty is to work out an age-scale. Accordingly Ballard set to work to remedy the defect and constructed a set of tests which he standardized according to an age-grade. The type of problems was the same as that of the Courtis tests: 28 problems in addition, 28 in subtraction, 28 in multiplication, and 28 in division. The Ballard tests are less difficult than the Courtis, however, as will be seen from the following examples :—

Addition.

64	35	82
16	29	63
31	40	9
98	78	14
22	51	23
75	47	65

Subtraction.

69152	80031	68703
48729	63175	37956

Multiplication.

273905	360197	591472
4	7	5

Division.

4 | 26930 7 | 66759 5 | 48175 6 | 44957

Allowance was made for the more simple character of the problems by reducing the time alloted to each operation, in this case three minutes being allowed for each one. One mark was allowed in the scoring for each answer absolutely correct. The following norms were obtained on the basis of the number of correct answers in a three minute performance :—

Age ...	9 years.	10 years.	11 years.	12 years.	13 years.	14 years.
Addition	3	4	5	6	7	8
Subtraction	2	3	4	5	6	7
Multiplication	1	3	4	5	6	7
Division	1	2	4	5	6	7

Ballard goes on to say that "if we mark the papers in another way and, instead of counting the number of sums right, count the number of operations right, we shall get a more exact score, for examples partly correct would score marks. By operations I mean processes of the kind tested. For instance, in the first addition example there are ten addition operations, in the first subtraction example five subtraction operations. For multiplication and division the corresponding numbers are six and four. The advantage of giving the norms in operation per minute, as in the following table, is that in applying a rough test any example may be set by a teacher, provided he makes a little allowance for the size of the sums, and the time taken in writing the figures and in passing from one sum to another."[1]

Number of Operations per Minute.

Age ...	9 years.	10 years.	11 years.	12 years.	13 years.	14 years.
Addition	12	16	20	24	27	30
Subtraction	4	6	8	10	12	13
Multiplication	4	7	10	12	14	16
Division	2	4	6	8	9	10

Reference has been made to one of the criticisms of the Courtis tests, viz., that it gives a grade-scale whereas an age-scale would be more satisfactory for purposes of comparison. Another criticism is that the Courtis tests are not diagnostic of the pupil's difficulties or of errors in teaching. They serve rather as measures of ability than as criteria for analyzing troubles. One of the groups of scales that has been constructed to obviate that criticism is the

[1] Ballard: Mental Tests, pp. 165, 166.

Woody Arithmetic Scales. The Woody scales were not primarily designed for diagnostic purposes, but have been found to serve that purpose rather well. The Courtis tests, as we observed, were constituted of problems of equal difficulty, but the Woody scales are made up of problems in a series arranged in an order of increasing difficulty. They are also designed to measure work in the four fundamental operations—addition, subtraction, multiplication and division. The addition scale covers problems with combinations in one, two, three and four column additions; examples with addends from 2 to 16; additions of simple fractions; addition of decimals; addition of United States currency; addition of denominate numbers; and addition of mixed numbers. Additions are expressed in two ways—by placing the digits in columns and by the plus sign. In that way the subject is tested in the entire range of problems calling for the operations of addition, the problems varying in possibility and in difficulty. What has been said of the addition scale applies also to the other scales in subtraction, multiplication and division.

We noted that the Woody scales have proved of value for the purposes of diagnosis. Wilson and Hoke have summarized[1] the following typical errors which were detected by means of the Woody scale for division:—

1. Ignorance of the multiplication tables, 30 per cent.
2. Using dividend as a whole, 14 per cent.
3. Confusion of multiplication and division, 14 per cent.
4. Remainder, 10 per cent.
5. Confusion of signs, 7 per cent.
6. Form of example strange, 5 per cent.
7. Carrying (either forgetting to carry or ignorance of what should be carried), 5 per cent.
8. Value of '0' 5 per cent.
9. Confusion of addition and multiplication, 5 per cent.
10. Confusion of dividend and divisor, 2 per cent.
11. Using some figure in dividend twice, 2 per cent.
12. Transposing answer, 1 per cent.

In a similar way a summary was made of the characteristic errors which recur in long division, as follows:—

1. The assumption that the first integer of the divisor may be used always as a trial divisor.
2. The trial-and-error method of finding the quotient.
3. Ignorance of the multiplication tables.
4. Carrying the wrong number when multiplying.
5. Borrowing in subtraction.
6. Ignorance of the value of the zero.
7. Forgetting to place integers in the quotient.

---

[1] See Wilson and Hoke: How to Measure, pp. 88, 89.

A good deal of valuable diagnostic work has been accomplished by various workers in the field on the basis of the tests which have been given. The limits of space do not permit me to go into the matter at any length, profitable as it might be. I can only refer the reader to the growing body of literature which deals with the problems. But for immediate consideration I would like again to quote Dr. Ballard who has given an immense amount of careful work to the questions and has summarized his recommendations[1] as follows:—

"1. That the tables, both addition and multiplication, be by some means or other fixed in the memory early in the arithmetic course.

"2. That the simultaneous repetition of the tables be superseded by individual learning, or better still, by their application to examples to be worked rapidly.

"3. That seriatim repetition be discarded after the structure of the tables is understood.

"4. That adding by tables be the final objective in practising addition, and that adding by units, or by partial groups, or throughout any roundabout device, be regarded as a habit of a lower order, to be abandoned as soon as habits of a higher order can be engendered.

"5. That speed of adding be insisted on as a means of pressing forward towards the higher habits.

"6. That the method of equal addition be universally taught as the practical method of working subtraction.

"7. That the method of decomposition be regarded, if taught at all, as a means of showing the correctness of the result arrived at by the usual method.

"8. That at least one pure practice lesson be given per week.

"9. That speed as well as accuracy be aimed at in the practice lesson.

"10. That the terminal examination in arithmetic contain at least one straightforward abstract sum.

"11. That each class be frequently practised in the work of the lower class.

"12. That means be adopted to secure the progress of each pupil at his own natural rate.

"13. That the blackboard be not used for setting out examples when text-books are available for that purpose; nor for working sums which could easily be worked by the majority of the class; nor for correcting errors due to mere carelessness. (The blackboard has, of course, its legitimate use for class and

[1] Ballard: Mental Tests, pp. 185, 186.

sectional teaching; it is only when it becomes a means of preventing individual effort that its use is open to objection).

"14. That the practice of copying in the exercise books examples worked on the board be discarded.

"15. That much of the responsibility of marking exercises be, with due reservations and precautions, delegated to the pupils."

II.—The Measurement of Ability to Read.

There are two kinds of reading to be measured: oral reading and silent reading. Tests have been devised for measuring each of these types, for the two types call for abilities which are quite disparate. In the case of oral reading it may be largely a mechanical art, and indeed ought to be developed to a certain degree from that standpoint. But in the case of silent reading the purpose is the acquiring of ideas.

A.—Oral Reading.

Binet thought that the fundamental process in reading was fluency by which he meant that there should be such pauses which are necessary for the elucidation of the passage read. So on that basis he constructed a reading scale which found a place in his *Barème d' Instruction.* Later investigators have put more emphasis upon the comprehension of a passage read than upon such matters as fluency, pronunciation, intonation, expression, etc. Comprehension is not an easy element to measure, but it involves a more elementary ability which is more readily measureable, viz., the ability to associate the appropriate sound images with the visual symbols that are presented by the printed page. And this ability is the fundamental factor in reading with comprehension.

To test this ability it is necessary to have a device which will overcome any tendency to anticipate, a certain amount of which is possible in the reading of a sensible passage. The discarding of sense material entirely and the putting together of words with no connection or association serves the purpose. Each word stands then by itself as a visual symbol which has to be translated into the appropriate sound. There is no possibility of grouping such as is done in phrases, but each word is a unit, and has to receive its full value. In this case it need scarcely be said that reading is a mechanical art. Some have criticized it very adversely on the ground that it amounts to nothing better than "barking at print." But those who have given the matter careful attention insist that it is basal to all reading, whether intelligent or otherwise. In estimating reading ability on this basis, the two factors of speed and accuracy are both taken into account. It is only fair that the tests should be so designed that they will not call for other factors such as are involved in the getting acquainted with new words.

So that none but common words must be used in an oral reading measure, for then only can we say that we are measuring the fundamental factor, the translation of the visual symbol into the associated sound. It is thus a test of visual perception, association, and appropriate articulation of sound. Dr. Ballard has given a test[1] which consists of 158 simple words which are printed in bold type and so arranged that they test oral reading as described. The subject is allowed one minute in which he should be able to read the 158 words, but no matter how proficient in reading he may be, he will find that it is not an easy task to complete the performance in the time required. The test is as follows :—

ONE MINUTE READING TEST.

is	me	on	at	by	so	us	an	it	or	be
to	as	he	of	in	go	up	am	if	no	we
my	ox	do	the	and	for	but	him			
are	can	he	dog	let	you	not	was			
out	try	see	mix	cat	now	boy	saw			
bit	met	top	run	man	pet	lot	get			
did	van	bad	red	cup	bee	lit	pin			
had	ran	pen	nut	big	old	yet	rob			
gun	leg	fun	lip	new	fog	has	sit			
sly	wig	mud	box	ink	sat	end	cut			
pay	fed	who	six	lad	wet	dry	cow			
his	peg	tin	say	eat	any	far	set	bud		
kid	pup	fox	ask	egg	cab	ill	use	jam		
act	toe	her	our	ten	arm	rock	gone	feel		
that	rich	till	long	flat	this	part	foot			
made	upon	came	mile	back	sand	time				
said	then	wall	into	were	done	walk				
much	loss	seem	went	with	come					

The above test was applied to the children in forty-nine schools on the basis of which the following norms were obtained :—

Age	6 yrs.	7 yrs.	8 yrs.	9 yrs.	10 yrs.	14 yrs.
Boys' Scores...	13	33	53	72	85	115
Girls' Scores...	15	38	58	76	88	112

B.—Silent Reading.

In the past the schools have given a great deal more attention to oral reading than to silent reading. The ability to read has too often been judged after the manner of an elocutionary contest. But actually silent reading is of far greater importance, because it

[1] See Ballard : Mental Tests, p. 136 for the test, and p. 139 for the table of norms.

is required in practically all subjects of the school curriculum, and because it is of much more use to the pupil after he leaves school. The truth is that the function of oral reading as a school subject is largely that of preparing the subject for silent reading. The rate with which a pupil is able to read silently is important as an indication of the manner in which his comprehensive ability functions. The criterion of measurement is thus qualitative as well as quantitative. We want to know not only how much a person can read silently in a given length of time, but how much of what he has read is comprehended.

Here in South India we are all familiar with the pernicious habit which some students tend to form of doing their preparatory reading orally. We are also familiar with the lament from many students about having too much work to do, more reading for their courses than they can hope to overtake. I need scarcely point out the intimate connection between these facts. The reason that many students are unable to cope with the volume of reading which their work demands is plainly that of faulty reading. Go into a room where a number of students are engaged in preparations, and the hum of voices is evidence that many of them are preparing by the method of oral reading. We need only experiment a very little to know that this means a great loss of time, for oral reading is a much tardier process than silent reading. I tried the experiment on one person who was able to read 385 words silently in a minute, but only 158 words orally from the same passage; another subject read 212 words silently and 54 words orally in the one minute, the oral reading in both cases being backwards so that it was purely a mechanical art. This wide difference serves to illustrate the loss of time in the case of students who read orally. If they would acquire the habit of reading silently and at the same time reading so as to comprehend the meaning of passages, they would be able to cover a much larger amount of work.

The method of conducting a silent reading test is fairly similar in all the tests of that kind. It consists of a number of passages which are printed on a test paper, and which increase slightly in comprehension difficulty. At the end of each passage a question is asked to answer which correctly the child must have comprehended the meaning of the passage. Sometimes a list of words is given, one of which is the correct answer and the child is asked either to underline or draw a circle around the one which answers the question correctly. At other times the instructions call for a more complicated response, which means a further drawing upon the comprehension of the subject. Each of the questions is carefully studied with reference to the responses made, and given a comprehension value on which the total score of the subject is obtained and which serves as a basis of comparison and standardization.

Dr. Ballard says that he knows of eight different silent reading tests which are in use in the United States to which he adds another of his own construction. Among the better known tests are the silent reading tests of Starch, Courtis, Monroe, Thorndike, and McCall. The Thorndike Visual Vocabulary Scale is based upon the understanding of the meaning of a paragraph requiring the ability to comprehend the meaning of the individual words which constitute the paragraph. The test is therefore so devised as to test that ability. By reproducing a portion we shall best appreciate Thorndike's method.

Thorndike's Reading Scale B. Word Knowledge or Visual Vocabulary—Series X.

Write the letter W under every word that means something about *war* or *fighting*.

Write the letter B under every word that means something about *business* or *money*.

Write the letters CHU under every word that means something about *church* or *religion*.

Write the letter R under every word like *father* or *wife* that means something about *relatives* or the *family*.

Write the letters COL under every word that means a *colour*.

Write the letter T under every word like *now* or *then* that means something to do with *time*.

Write the letter D under every word like *here* or *north* that means something about *distance* or *direction* or *location*.

Write the letter N under every word like *ten* or *much* that means something about *number* or *quantity*.

4·0 camp, flag, west, mother, two, general, green, troops, south, fort.

4·5 gray, cousin, pink, uncle, yellow, hour, pay, aunt, early, commander.

5·0 marriage, defeat, many, afternoon, guard, buy, captive, military, relation, late.

6·0 hymn, defend, across, merchant, noon, forty, conquer, dagger, profit, Tuesday.

There are eight lists similar to that reproduced and in each case there is graduated difficulty. The pupil's score is reckoned as the value of the most difficult line in which he succeeds in marking eight out of the ten words correctly.

In the Thorndike-McCall Reading Scale there are 35 passages given which the subjects are instructed to read and at the conclusion of each they are asked to respond to certain questions. McCall reproduces an easy and a difficult portion from the scale in his book, *How to Measure in Education* together with the questions which are based upon the passages. They are as follows:

1. Nell's mother went to the store on Water Street to buy ten pounds of sugar, a dozen eggs and a bag of salt. She paid a

dollar in all. Nell and Joe went with her. On the way home on Pine Street, they saw a fire-engine with three horses.

i. Was the salt in a box or a bag or a can or a dish? . . .
ii. How many eggs did she buy? . . . . . .
iii. What did the children see on Pine Street? . . . . .
iv. What street was the store on? . . . . . .

31. COLERIDGE.

I see thee pine like her in golden story
Who, when the web—so frail, so transitory,
The gates thrown open—saw the sunbeams play
With only a web 'tween her and summer's glory
Who, when the web—so frail, so transitory,
It broke before her breath—had fallen away,
Saw other webs and others rise for aye,
Which kept her prisoned till her hair was hoary.
Those songs half-sung that yet were all divine—
That woke Romance, the queen, to reign afresh—
Had been but preludes from that lyre of thine,
Could thy rare spirit's wings have pierced the mesh
Spun by the wizard who compels the flesh,
But lets the poet see how heav'n can shine.

xxx. Who acted like a spider? . . . . . .
xxxi. Who or what is compared with a woman? . . . . . .
xxxii. Copy the first word of the line which implies there has not been a continuous stream of such songs? . . . . . .
xxxiii. Complete the following with one word only:—
"Those songs" really means those. . . . . .

The results of the test are studied very minutely with reference to many factors, but chiefly with the purpose of enabling the teacher to guide the student in remedying his defects which will be diagnosed by means of the test.

One group of the silent reading tests which has been used the most is that known as The Standardized Silent Reading Tests which were devised by Prof. Walter S. Monroe. These tests are arranged in three groups, the first for grades 3, 4 and 5; the second for grades 6, 7 and 8; and the third for grades 9, 10, 11 and 12. Each exercise is scored in two ways. It is given a rate value which indicates the number of words read per minute in careful reading, and a comprehension value which represents the scoring of the child's ability in understanding what he has read. The first test has 15 exercises with a total rate value of 123 and a total comprehension value of 29·5; the second test has 13 exercises with a total rate value of 162 and comprehension value of 44·7; the third test consists of 12 exercises having a total rate value of 145 and a comprehension value of 72·5. The Monroe tests have been very largely used so that a great deal has been done in standardizing

them, though additional data is constantly modifying the medians to a slight extent. The standards for the middle of each year for the different grades are given by Wilson and Hoke as follows:—

Grade.	III	IV	V	VI	VII	VIII	IX	X	XI	XII
Rate	52	73	89	88	99	106	87	81	88	89
Comprehension ...	7·2	13	19	20	23	26·4	25	25	26·4	27·2

A few samples of the Monroe Standardized Silent Reading Tests will serve to indicate the type that is employed. The Kansas Silent Reading Tests are much like the Monroe tests. They were devised by Dr. F. J. Kelley and their best features have been incorporated in the Monroe Tests.

Quoted from Test No. I.

No. 1.

Rate value.		Comprehension value.
9	The little red hen was in the farmyard with her chickens, when she found a grain of wheat. "Who will plant this wheat?" she said. Draw a line under the word which tells where the little red hen was. barn chicken-house feed bin farmyard	1·1

No. 7.

11	The door opened and in came the dog. The mice jumped off the table and ran into the hole in the floor. The poor little country mouse was so frightened! What frightened the mice? Draw a line under the word that tells what it was that frightened the mice. boy woman cat trap man dog wind	1·7

No. 14.

10	On the ground the apples lie. In piles like jewels shining. And redder still on old stone walls Are leaves of woodbine climbing. What time of year is pictured? If spring, draw a line under "winter." If not, draw a line around the right season. spring summer fall winter	2·8

Quoted from Test No. II.

No. 5.

Rate value.		Comprehension value.
11	The caravan, stretched out upon the desert, was very picturesque; in motion, however, it was like a lazy serpent. By and by its stubborn dragging became intolerably irksome to Balthasar, patient as he was. Place a line under the word which tells in what respect the caravan resembled a serpent. colour length motion size	3·2

No. 8.

12	Judah walked in the pilot's quarter. So absorbed was he in thought that he scarcely noticed the shores of the river which were surpassingly beautiful, with orchards of fruits and vines. If he is interested in the beauties around him, put a line under beautiful; if these beauties have no interest for him, put a line under shadow. beautiful shadow	3·7

Quoted from Test No. III.

No. 1.

9	Smoke is lighter than air. Too much smoke in the atmosphere will suffocate a person. John is in a smoke-filled room and cannot get out. If he should stand, underline smoke. If he should lie on the floor, underline air. smoke room air atmosphere	3·5

No. 6.

15	The expressionless uniform twenty houses, all to be knocked at and rung at in the same form, all approachable by the same dull steps, all fenced off by the same pattern of railing, all with the same fire escapes, and everything without exception to be taken at the same high valuation. After reading the above paragraph, underline the word that tells what you think would be the general effect of the street. variety attractiveness monotony beauty	5·4

III.—The Measurement of Ability in Spelling.

The ability to spell correctly is called into play when a person is writing, but not in conversation. It is needed in such social processes as writing letters, business notes, articles, and so forth. The psychological process involved in spelling is not by any means simple, as is evident from the critical examination of the processes by Prof. Leta S. Hollingworth[1] to which reference has already been made. The process involves the formation of a series of associations or "bonds" which she describes as follows :—

"(1) An object, act, quality, relation, etc., is 'bound' to a certain sound, which has often been repeated while the object is pointed at, act performed, etc. In order that the bond may become definitely established, it is necessary (*a*) that the individual should be able to identify in consciousness the object, act, quality, etc., and (*b*) that he should be able to recollect the particular vocal sounds which have been associated therewith.

"(2) The sound (word) becomes 'bound' with performance of the very complex muscular act necessary for articulating it.

"(3) Certain printed or written symbols, arbitrarily chosen, visually representing sound combinations, become 'bound' (*a*) with the recognized objects, acts, etc., and (*b*) with their vocal representatives, so that when these symbols are presented to sight, the word can be uttered by the perceiving individual. This is what we call ability 'to read' the word.

"(4) The separate symbols (letters) become associated with each other in the proper sequence, and have the effect of calling each other up to consciousness in the prescribed order. When this has taken place we say that the individual can *spell orally.*

"(5) The child by a slow, voluntary process 'binds' the visual perception of the separate letters with the muscular movements of arm, hand, and fingers necessary to *copy* the word.

"(6) The child 'binds' the representatives in consciousness of the visual symbols with the motor responses necessary to produce the written word spontaneously, at pleasure."

In selecting words which shall be used in testing ability in spelling, there are certain criteria which are to be borne in mind. The chief of these are frequency, difficulty, number, and administration. We want to know what are the most commonly used in the language, ability to spell in which is being tested. We want to know something about the relative difficulty of words. We need to know how extensive to make the test—how many words should be included. And we need to know the best method for administering the test for the most satisfactory results.

---

[1] Hollingworth, L. S., and Winford, C. A.: The Psychology of Special Disability in Spelling, in the Teachers' College Record, Columbia University, March 1919.

Measuring the ability to spell by standardized tests is, so far as I have been able to learn, confined to the English language. So that what we are able to conclude is in regard to spelling English words only.

Some most labourious investigations have been carried on by those who have tried to work out a scale for the measurement of ability in spelling. Dr. W. F. Jones of the University of South Dakota spent eight years conducting an investigation which covered four states. He arranged for the writing of 75,000 themes written by 1,050 pupils on a variety of subjects sufficiently large to bring into play their entire vocabularies. The total data covered over 15,000,000 words. The number of compositions written by the various pupils varied from 56 to 105. Jones found that there were only 4,532 different words which had been used by all of the pupils. The largest single vocabulary was that of an eighth-grade girl and included 2,812 different words.

Another celebrated investigation was that conducted by Dr. Leonard P. Ayres of the Russell Sage Foundation, New York City. The data for his scale was computed from an aggregate of 1,400,000 spellings by 70,000 pupils in the schools of 84 different cities throughout the United States. In addition to the material which he collected from the compositions of school-children, he also used letters, newspapers, standard literature, etc., in order to discover what were the most frequently used words. Ayres is in entire agreement with Jones as to the fundamental conclusions, viz., that the writing vocabulary of the majority of persons is both small in compass and made up of simple words. Ayres found that the vast majority of words which we use in practical life, excepting technical and scientific words total only about 1,000. He discovered that there are 50 words which are used so frequently that they comprise about 50 per cent of our vocabularies in English.

On the basis of the investigations Ayres constructed his scale, dividing the words into 26 groups, lettered from "A" to "Z". In group "A" there are two words—"me" and "do"—which were spelled correctly by 99 per cent of second grade pupils; while at the other end of the scale in Group "Z" there are three words—"judgment",[1] "recommend" and "allege"—which were spelled correctly by only 50 per cent of eighth grade pupils. All of the words in each column are of approximately equal spelling difficulty. At the top of each column is indicated the average per cent of the words spelled by each grade from 50 per cent and upwards.

[1] It would be interesting to know whether Ayres marked the preferable spelling "judgement" as wrong. Certainly a spelling which has the imprimatur of Oxford University, should not be discredited. The obvious reason is that "g" not followed by "e" or "i" is usually pronounced as hard "g" whereas in judgement the "g" is soft.

No record is made of averages below 50 per cent. Blank spaces to the left indicate that the children of those particular grades spell all of the words in those groups correctly. For example children of the eighth grade are averaged at 100 per cent for all the columns "A" to "N" inclusive; 99 per cent for "O"; 98 per cent for "P"; 96 per cent for "Q"; 94 per cent for "R"; 92 per cent for "S"; 88 per cent for "T"; 84 per cent for "U"; 79 per cent for "V"; 73 per cent for "W"; 66 per cent for "X"; 58 per cent for "Y"; 50 per cent for "Z". The words in column "N" may be quoted here as those of median difficulty, but the whole scale should be studied by all those who have to teach spelling in English.

Column "N": except, aunt, capture, wrote, else, bridge, offer, suffer, built, centre, front, rule, carry, chain, death, learn, wonder, tire, pair, check, heard, inspect, itself, always, something, write, expect, need, thus, woman, young, fair, dollar, evening, plan, broke, feel, sure, least, sorry, press, God, teacher, November, subject, history, April, cause, study, himself, matter, use, thought, person, nor, January, mean, vote, court, copy, act, been, yesterday, among, question, doctor, hear, size, December, dozen, tax, number, October, reason, fifth.—75 words.

There are several other lists which have been arranged, some of which are extensions or imitations of the Ayres' scale, but none of which have been tried out so thoroughly. Dr. Buckingham has extended the Ayres scale by six steps with 505 more words, making it useful for upper grades and high school subjects, but the additions are not as fundamentally important words as the original scale. Buckingham has also carried on an investigation in regard to the relative difficulty of spelling words. Wilson and Hoke, writing in 1921, say that Buckingham is working on the preparation of a list of 1,000 words arranged in order of difficulty. The Iowa Spelling Scale is a list of 2,977 words so arranged as to imitate the Ayres Scale. The Rice Test prepared by Dr. J. M. Rice consists of three tests, the first of which is a list of 50 words, the second a composition passage containing 50 other words which he wished to give, and the third test was a composition test based upon a picture in which case the pupils were required to select their own words and spell them. The Starch Test is a list of 600 words divided into six parts of 100 words each. The selection of words was at random from a dictionary—the first defined word on each even-numbered page of the 1910 edition of the New International Dictionary being chosen, with the exception that proper names, technical words and obsolete words were discarded from the list. The test is unsuited to the lower grades as it contains

many difficult words. The Boston Schools have prepared a list of their own, and the Normal School at Chico, California, another.

Mr. Cyril Burt has drawn up a list for use in English schools which Ballard agrees is the best for the purpose. It is considerably shorter than the Ayres Scale, and is arranged on the age-scale plan, from 5 to 14, the findings being that approximately 50 per cent of each age will spell the words correctly which he has assigned to that age. The following is his list: a total of 100 words, 10 to the year.

Burt's Graded Spelling Test—

Age.

5. a it cat to and the on up if box.
6. run bad but will pin cap men got to-day this.
7. table even fill black only coming sorry done lesson smoke.
8. money sugar number bright ticket speak yellow doctor sometimes already.
9. rough raise scrape manner publish touch feel answer several towel.
10. surface pleasant saucer whistle razor vegetable improvement succeed beginning accident.
11. decide business carriage rogue receive usually pigeon practical quantity knuckle.
12. distinguish experience disease sympathy illegal responsible agriculture intelligent artificial peculiar.
13. luxurious conceited leopard barbarian occasion disappoint necessary treacherous descendant precipice.
14. virtuous memoranda glazier circuit decision mosquito promiscuous assassinate embarrassing tyrannous.

A most interesting and important study is that of misspelling. Several investigators have experimented in this field, including Dr. Leta S. Hollingworth, C. A. Winford, S. A. Courtis, Ayres, Jones, A. W. Kallom, F. N. Freeman, and others. The most interesting result is that of Dr. Jones who, on the basis of his extended investigation of the spelling of pupils in composition, prepared a list of the words which were misspelled the most frequently. This list is known as "The One Hundred Spelling Demons of the English Language." Appended is the list.

which 321	meant 247	minute 210	often 185
their 316	just 245	busy 209	writing 184
there 296	many 245	two 208	doctor 182
separate 283	too 243	much 206	very 182
hear 280	Tuesday 242	enough 206	though 181
here 278	knew 237	seems 205	among 179
said 275	lose 236	none 203	sure 179

been 273	week 235	does 203	tonight 174
says 273	can't 234	easy 202	forty 172
they 271	grammar 234	would 200	since 172
some 270	whole 231	whether 200	once 170
any 268	wear 230	loose 198	raise 169
Wednesday 266	every 228	could 196	trouble 168
done 263	instead 228	ready 196	choose 168
know 263	built 225	beginning 195	colour 167
read ("red") 261	blue 224	heard 195	dear 166
piece 260	shoes 224	country 194	truly 166
don't 258	won't 221	business 194	early 166
break 257	wrote 220	ache 192	used 165
tear 255	cough 217	answer 191	friend 164
February 255	where 216	making 190	again 164
laid 252	write 216	always 188	hoarse 162
straight 251	buy 212	hour 187	guess 162
through 250	believe 212	tired 187	women 161
half 250	coming 212	sugar 185	having 158

The difficulties in connection with misspelling have been a matter of much thought and several methods of correcting them have been suggested. One good way is to interest the pupils in making lists of their own misspelled words whereby they will develop a spelling conscience and improve their spelling habits. Various methods may be used by the educator to interest the pupil so that it will not appear a mere drudgery. It can be done frequently by means of games, Monroe suggesting[1] the following as available for the purpose:

1. Syllable game.
2. Jumbled-letter game.
3. Initial game.
4. Rhyming game.
5. Derivative game.
6. Definition game.
7. Linked-word game.
8. Missing-word game.
9. Composition game.

One of the questions to be considered in testing spelling ability is that of the rate at which the dictation should be given. This is a matter which also concerns ability in handwriting. Professor Freeman has made an investigation of the rates, and has standardized the rates of handwriting for the various grades to be observed in dictation. He found that pupils wrote the following number of letters per minute; second grade, 36 letters; third grade, 48 letters fourth grade, 56 letters; fifth grade, 65 letters; sixth grade, 72 letters; seventh grade, 80 letters; eighth grade, 90 letters. As this

[1] Measuring the Results of Teaching, p. 194.

is the rate of handwriting, dictation should be a little bit slower to allow for the translation of the sound into the visual image before writing. Probably ten per cent additional time would be about right. On this basis he suggests the following as the number of seconds to be allowed per second for the various grades:

Grade	Seconds per letter.
II	1·83
III	1·38
IV	1·18
V	1·01
VI	·92
VII	·83
VIII	·73

If sentences contain more than thirty or forty letters the dictation should be in sections rather than all at once. Furthermore all pupils do not write at the same rate of speed so that provision must be made for the slow writers, especially since the test is one of spelling. It is generally recognized that the tests are better administered when the words are embodied in sentences than when they are dictated in columns.

IV.—Measuring Ability in Handwriting.

The measurement of ability in handwriting is plainly a more difficult task than measuring ability in spelling or arithmetical processes, because of the fact that we cannot have as fixed a standard. There is a great deal more scope for subjectivity in the judgements of teachers as to what constitutes good and what bad handwriting. In spelling we have definite objective standards. Sometimes there are alternative spellings which are correct, but outside of that scope we know definitely when a word is misspelled by reference to the standard which is preserved for us in the dictionary. In handwriting there are many styles and many variations of judgement and even the scales that have been attempted illustrate this factor of subjectivity.

What are the factors of which we must take cognizance in the measurement of handwriting? The answer is: two,—quality and speed. Speed is not difficult to determine with reference to a standard. It is judged by counting the number of letters written during a given period and reducing to a basis of so many per minute. Quality is measured by securing specimens of the pupil's handwriting and comparing it with the specimens in a handwriting scale.

But how is the scale constructed? At first blush one might suspect that its construction is totally a matter of opinion. But as a matter of fact it is carefully done with reference to a number of factors. Drs. F. N. Freeman and Truman Gray have each of them

made valuable contributions to the analysis of the factors which must be observed. Dr. Freeman's analysis' is one of the defects in writing and of their causes. It is as follows:

	Defect.	*Causes.*
1.	Too much slant	(1) Writing arm too near body.
		(2) Thumb too stiff.
		(3) Point of nib too far from fingers.
		(4) Paper in wrong position.
		(5) Stroke in wrong position.
2.	Writing too straight.	(1) Arm too far from body.
		(2) Fingers too near nib.
		(3) Index finger alone guiding pen.
		(4) Incorrect position of paper.
3.	Writing too heavy.	(1) Index finger pressing too heavily.
		(2) Using wrong pen.
		(3) Penholder too small diameter.
4.	Writing too light.	(1) Pen held too obliquely or too straight.
		(2) Eyelet of pen turned side.
		(3) Penholder too large diameter.
5.	Writing too angular.	(1) Thumb too stiff.
		(2) Penholder too lightly held.
		(3) Movement too slow.
6.	Writing too irregular.	(1) Lack of freedom of movement.
		(2) Movement of hand too slow.
		(3) Pen gripping.
7.	Spacing too wide.	(1) Pen progresses too fast to the right.
		(2) Too much lateral movement.

Dr. Gray has put his analysis into the form of a score card and it has the advantage over that of Freeman that the analysis is positive rather than negative. Moreover the analysis is more complete if anything than that of Freeman. The card calls for marking on a percentage basis the 100 per cent being divided among five main factors.

Factors.	Percentage of marks.	Factors.	Percentage of marks.
1. Heaviness	3	5. Spacing of lines ...	9
2. Slant	5	Uniformity.	
Uniformity.		Too close.	
Mixed.		Too far apart.	
3. Size	7		
Uniformity.		6. Spacing of words ...	11
Too large.		Uniformity.	
Too small.		Too close.	
4. Alignment	8	Too far apart.	

[1] Freeman, F. N.: *The Teaching of Handwriting*, p. 72.

Factors.	Percentage of marks.
7. Spacing of letters ...	18
Uniformity.	
Too close.	
Too far apart.	
8. Neatness	13
Blotches.	
Carelessness.	
9. Formation of letters ...	26
General form	8
Smoothness	6
Letters not closed	5
Parts omitted	5
Parts added	2
Total score ...	100

In giving a test for handwriting, as in the case of testing other abilities, there are certain rules which ought to be observed for the obtaining of the best results. Wilson and Hoke have summarized them into the following points, as follows :—

1. It is necessary to have a simple, easily understood copy which even second-grade pupils can comprehend.
2. Pupils should be required to memorize the copy before beginning the test, since it is to be a test of handwriting, measuring speed as well as quality.
3. The time must be accurately determined and standardized for purposes of comparison.
4. All preparations must be complete before the signal to start is given, so that all may have an equal opportunity.
5. The teacher must give the directions simply and explicitly.
6. The pupils may be required to count the number of letters written, and save the teacher's time to that extent.

Several scales for judging handwriting have been devised. One of the most important is that of Ayres, which consists of twenty-four samples of writing, eight each of the vertical, semi-slant and full slant styles. In each of the three styles there is a grade for 20, 30, 40, 50, 60, 70, 80 and 90. Another scale is that of Thorndike, and is based on the three characteristics—beauty, legibility and general merit, the degree of these three characteristics represented in the specimens of the scale having been determined by the consensus of opinion of competent judges. The scoring in the Thorndike scale is from 4 to 18, and one or more specimens are furnished for each degree of quality represented. It has been found possible to compare the resultants of these two scales by multiplying the score in the Thorndike scale by 6·7 and subtracting 20 from the product in each case. Freeman's scale really comprises five scales, one to measure each of the following characteristics:—uniformity of slant, uniformity of alignment quality of line, letter formation, and spacing. These are printed in the form of a chart, each scale constituting a division. There are other scales in existence but those mentioned are representative. Experiment have been conducted in several cities and some

of the States and the results have been used in obtaining medians Records may be consulted in the books of Monroe, and Wilson and Hoke to which references have already been made.

V.—Measuring Ability in Composition.

There is scarcely any subject of more practical importance in ordinary life than composition, and yet there is none more baffling to the inventor of scales of measurement. There is no subject which commands so much of the instructor's time and gives him so much worry. In many cases he feels that it is time lost, for he has no assurance that the pupils are going to give any attention to his red ink notations, made in an effort to help the pupils to correct their errors and improve their style. Here in South India the difficulty is augmented on account of the fact that we have to do with more than one language, and in each case we have to measure ability in composition.

In the United States a number of scales have been devised for the marking of English composition, including those of Hillegas, the Nassau County Supplement to the Hillegas Scale, the Thorndike Extension of the Hillegas Scale, the Willing Composition Scale, the Gray Composition Scale, the Harvard-Newton Scales for the Measurement of English Composition, and Breed and Frostic's Scale for Measuring the General Merit of English Composition.

The method of measurement is much the same as that used in measuring handwriting. A number of themes are arranged in order of merit, and are taken as specimens with which to compare the production of the pupil. In the Willing scale, for example, a number of compositions on the topic, "An Exciting Experience" are arranged on the evaluation of 20 to 90 by tens. These marks are more or less arbitrary, 20, e.g., signifying 15 to 24·9, 30 signifying 25 to 34·9, and so on. Under the grade marked 20 the number of mistakes in spelling, punctuation and syntax per hundred words is placed at 30, in the grade marked 30 at 23, in the grade of 40 at 17, in the grade of 50 at 14, in the grade of 60 at 11, in the grade of 70 at 8, in the grade of 80 at 5, and in the grade of 90 at zero.

The Hillegas scale which was the first in the field consists of ten compositions arranged in order of merit, the marks given ranging from 0 to 9·3. The difficulty with the scale is that there is so much variation, three being artificial productions, five written by high-school students and two by college freshmen. They were all on different themes, and the length varies greatly. In the Thorndike extension of this scale only a few of the original composition specimens have been retained, whereas the number of specimens has been increased to twenty-nine, representing fifteen

degrees of merit, the values ranging from zero to 95. This is the scale which is probably in most common use.

The tests so far devised are obviously devised as measures of ability in written composition only. As yet no one has constructed any scale for oral composition, which would be a still greater problem. There are difficulties enough in the work of measuring written work, and no one scale is above criticism. Ballard, e.g., levels the criticism of insularity, and thinks that Thorndike's examples are unsuited as a scale to be used in English schools. It is possible that another scale will be necessary for use in India, different again from one which be applicable to either American or English conditions. Yet in the interests of standardization it ought to be possible in time to devise a scale for measuring ability in English Composition for subjects in any part of the world.

Mention has been made of the measurement of attainment and progress in five subjects only. These five have been selected for the very simple reason that more has been done to devise scales for measuring these abilities than other subjects. Still some work has been done in other fields. High school and college subjects admit of a greater variety of correct performance than the public school grades, and hence the task is more difficult. Yet tests have been constructed and a considerable amount has been done in standardizing them in Algebra (Monroe, Hotz, and Rugg and Clark), Geometry (Stockard and Bell), Physics (Starch), Latin (Henmon and Starch), French (Starch and Henmon). Ancient History (Sackett), Commercial Subjects (Sherwin Cody), Geography (Hahn-Lackey, Buckingham, Starch, Witham, and Branom and Reavis), and Practical Ability (Ballard, Burt, McDougall, etc). The volume of work that must be done to standardize the measurements in all of these subjects is overwhelming. The hope is in the small army of educational psychologists who are giving themselves to the work.

CHAPTER IX.

THE STATISTICAL STUDY OF RESULTS.

The application of the art of measurement to the study of mental abilities carries with it the necessity for using certain quantitative devices. It can scarcely be said however that the measuring of mental abilities is equivalent to the reduction of the qualitative to the quantitative. We are quite ready to admit that, in dealing with psychological data, many qualitative factors appear, but our purpose is to find out as nearly as possible the extent to which they are present. It is a comparative procedure. Standards are set up in the interests of comparison. Our ultimate concern is the comparison of an ability in one person with the same in another, or in one group with another. The standard which we set up is some artificial device, such as an intelligence quotient or an educational quotient, which serves as a sort of medium of comparison.

The science which is particularly concerned with matters of this kind, and on which we may call for assistance in making our measurements is the science of statistics. Statistics concerns itself with a systematic collocation of numerical data in relation to the enumeration of groups or to the ratios of quantities associated with such groups which have been obtained by the method of enumeration. In mental measurement we are concerned with measurements of intelligence and of attainment on the basis of which we desire to make certain distributions of scores and to make comparisons between the groups. So that we have in statistics the precise mechanism which we need to complete our study and interpret our results. Statistics is a branch of the mathematical disciplines, and includes problems which call for skilled mathematical technique. At the same time there are problems of a less complicated nature which concern us in this science in which statistics may help us without our needing to take an honours course in mathematics.

The immediate purpose which concerns us is the reduction of mental measurement to a science. Science is an essentially mechanical technique, and deals with its data in such a way as to make the future as mathematically calculable as possible. One of it chief characteristics is accuracy. It attempts to overcome all tendencies to guess work and haphazard conclusions based on insufficient data. The difficulty with educational methods in the past was precisely its lack of a scientific technique. Now a scientific study of educational problems involves in the first place a systematic observation of educational conditions so to collect the necessary facts and record the observations upon which any generalizations must be determined. The scientific method of to-day is the inductive method, and no induction is valid which has not observed

and collocated a sufficient number of facts. In the second place, if education is to be scientific it must devise criteria of measurements. There is something the matter with an educational criterion that pronounces a pupil fair whose chronological age is fourteen, and is in a class the average age of the members of which is ten. There has long since been agreement upon what we mean by a "yard" in measuring cloth, or a "degree" in measuring temperature, or a "rupee" in measuring market value. The scientific method in education voices the demand that we should reach some agreement when we are talking about intelligence, or ability in arithmetic, or in spelling, or in motor skill, or about any other psychological facts. The scientific method makes use of data which have been gathered from all sources. It is only a few years—perhaps twenty—since education began to make use of cognate facts obtained by the biological and physical sciences, and in particular of the statistical method. Scientific method to-day lays a great deal of stress upon experimentation. It is the method of the laboratory. As far as the educationalist is concerned, if his work is to take the character of science he must regard the school-room in a sense as a laboratory, always remembering, of course, that the centre of interest is the child, yet for the sake of the child's normal development being willing to experiment along any line that promises to yield fruitful results.

We noted that a prime necessity for scientific study is the systematic observation of facts. That is characteristic of mental measuring. In fairness to the subject, the experimenter tries to reserve his conclusions until he has summoned to his aid all the available facts that are relevant. Before giving a test to a child or to a class, it is usual to consult the records for any data which will give light on the child's environment, his past history, his physical condition, his school progress, his habits, his temperamental characteristics, his age, the average age of members of his class, his standing in the class as indicated by the school examinations, the judgement of his teacher, and any other facts that are obtainable. When we are dealing with human personalities, the greatest values in the world, we cannot afford to neglect any data available in making our judgements. Moreover the greater the number of facts which we can collect in regard to the individuals or groups of persons, through the channels of psychological tests, the more scientific will be our conclusions. The criticism of the original Binet tests was that they were too narrow in scope, and the Stanford revisers have done well to broaden them by the addition of more tests. But, if we accept the theory of intelligence as made up of a number of abilities, and we must accept of attainment as including a number of specific abilities, it is plainly impossible to reach sound conclusions on meagre data. For the results which we reach in regard to any particular ability cannot be considered as holding in

regard to any other ability. Experiments in testing arithmetical ability have led investigators to conclude that there is no general arithmetical ability even, but that the various processes call for the functioning of different abilities. This involves the necessity for wide-scoped observation for any scientific conclusion as to the abilities of any individual.

And when it comes to building up data sufficient for the reaching of averages and standards it is again apparent that the observations must be wide if they are to lead to valid conclusions. In the case of group factors, they admit of a variety of influences as broad as the individual. There are such factors as social strata, racial characteristics, physical differences, caste influences, school advantages, sanitary conditions, religious inheritances, and any other group sanctions. If educational medians and scales are to be valid everywhere and among all classes, it means the amassing of an immense amount of data on the basis of which results are calculated. We find men of one group complaining of standards and scales set up by workers among other groups. Some of the Binet-Simon tests were said to be all right for French children but unsuited to American and English children. And some of the American tests are said to be all right for children of the United States, but not for any other country. And now that work is beginning in India, workers are beginning to find certain defects in the case of existing tests because they do not suit Indian communities. Only an immense amount of observation and collection of data will be able to solve the problem of whether or not it will be possible to get a scale of tests that will be suited to all communities. And if such is impossible, it will mean much labour to ascertain how tests can be adapted so that the standards will not be spoiled.

I.—Devising a Scale.

The construction of a scale calls for the operation of statistical methods. In other words, it is first necessary to test a test with children, before using it as a test for children. We have already observed that the Binet scale was constructed on that principle. If he found a test were passed by from 65 to 75 per cent of children of a certain chronological age that appeared to be normal, he took the test as valid for that age mentality. For example, if he tested 100 ten-year-old children with a certain test and found that from 65 to 75 of them succeeded, he included it in his scale as a ten-year-old test. Binet's first scale was constructed after testing 200 children. It is no wonder that he had to revise it. In the very nature of the case the more there are tested, the more likelihood there is of the percentages shifting, so that the places given by Binet to tests in his scale were sometimes shifted, as we have seen, as much as three years. Obviously nothing less than a very large number of

cases tested could yield data sufficient to be sure that the averages would not be materially altered by subsequent investigation.

The same method was used by the American Army psychologists in preparing their scales for the examination of enlisted men. To begin with an examination was made of the availabe tests, and a committee of experts sifted and selected these and constructed a scale for preliminary examination. But before testing men with the test, they tested the tests with men. In four of the training camps 80,000 men were examined, and in high and elementary schools 7,000 students were also examined in what has been called "the official trial of the method." Then before the tests were put to use as scales of measurement, the data assembled from the official trials were subjected to meticulous statistical treatment by a core of experts. On this basis again the psychologists of the various camps and members of the original committee spent two months in studying the results and revising the methods, the final outcome being the result of all that labour. According to Yoakum and Yerkes: "The validity of the tests as measures of intelligence was checked against every available criterion, including officer rating of men, army rank as an outcome of the survival of the fittest, other kinds of intelligence scales, professional success, and ability to learn as evidenced by school standing . . . The influence of literacy, repetition of the test, physical condition of the examinee, and the personal equation of the examiner have all been carefully considered."[1]

McCall mentions three characteristics by which to test a test, and the same might be applied to a scale of tests. These are validity, reliability and objectivity. Then he quotes the National Association of Directors of Educational Research of the United States as defining validity in terms of "the correspondence between the ability measured by the test and ability as otherwise objectively defined and measured. When a test really measures what it purports to measure and consistently measures this same something throughout the entire range of the test it is a valid test.".[2] A valid test is one which reproduces some process which is fundamental to life. A test which is intended to measure the ability of a person in spelling, and yet does not call for the spelling of the words which the person ordinarily uses and with which he is familiar, would not be valid. A test which is intended to measure arithmetical ability must stand the same test. Imagine a commercial man who is familiar with commercial arithmetic having his arithmetical ability tested by quadratic equations. Again a valid vocational test must also be related vitally to the vocation and to the subject. College graduates sometimes make miserable failures when put at certain occupations in spite of brilliant college records, and such

[1] Army Mental Tests, p. 9.

[2] How to Measure in Education, p. 195.

chagrin might be avoided by the application of a valid vocational test without the necessity of spending weeks or months perhaps in the vocation itself.

The problem of determining validity involves also correlation which is a distinctly statistical study. We shall return to it presently. Let it suffice to point out here that correlation with other measures is one of the best indications of the validity of a test or of a scale of tests.

The second characteristic of a test is reliability which McCall defines as "the amount of agreement between results secured from two or more applications of a test to the same pupils by the same examiner. Perfect reliability obtains when an identical examiner applies two identical or exactly duplicate tests according to an identical procedure to identical pupils."[1]. The precision of the language which McCall uses here is an indication by contrast of the many possibilities through which unreliability may arise. External conditions may affect either the examiner or the examinee or both. The nature of the test may be such as to induce similar effects. For example if the instructions are not explicit such may very easily arise, and the greatest amount of care should be taken to see that instructions are explicit and incapable of two interpretations. Another factor which must never be forgotten is that the psycho-physical organism is always in process of change, and any test or scale that is constructed on the understanding of the organism as something static is doomed to failure. The only safe way of testing the reliability of a test is to apply it to the same pupil or to the same pupils on two or more occasions, and to compute the correlation between the various performances. If the correlation be one, then we have the best evidence of good reliability; if it be zero, we have evidence of absolute unreliability.

McCall's third characteristic of a good test is objectivity. A test may be described as objective when two or more applications of the same test to the same pupils by different examiners yield identical results. If no agreement be reached by two or more examiners on the results of a test, it is perfectly subjective. Objectivity and reliability are both relative factors, so that neither can be expected to give us an absolute criterion. Some tests lend themselves to objectivity much more than do others. They do not admit so much of the examiners' differences, nor of the changes in the subjects. There is, of course, an intimate relation between reliability and objectivity, which McCall has expressed in the form of an equation, thus:

"Objectivity = reliability—personal equation." [2]

When a test satisfies the conditions of validity, reliability and objectivity, there still remains the task of fixing its place in the

---

[1] Op. cit., p. 307.

[2] *Ibid*, p. 313.

scale. Scaling the test is of the great importance because the child is the center of interest. There is no other interest in devising and applying tests than the discovery of differences between children and groups with a view to giving all an opportunity for normal development up to the maximum. This involves the matter of the distribution of the scores and their treatment by statistical methods so that some common denominator can be obtained. Several methods are in vogue for the scaling of tests of which the commoner are the grade-scale, the age-scale, the percentile scale, the product scale, and the T-scale of McCall and Thorndike.

The grade scale proceeds by a grade variability unit. A grade scale for any grade demands some measure of the variability of the performance of pupils. Such units as the Standard Deviation (S. D.) and Probable Error (P. E.) are used for that purpose. It is usual to take the median as a central point from which deviation is measured. Standard deviation is determined by taking the square root of the sum of the squares of the deviations from the arithmetical mean or average. Probable error is an expression which has survived the days when deviations were considered to be errors, and when the "curve of error" was another expression for the normal curve of distribution. It is obtained by multiplying the standard deviation by ·675 or more accurately · 67449, where the curve of distribution is normal.

Formula : P.E. = S. D. × ·67449.

Woody, in his *Measurement of Some Achievements in Arithmetic* gives the details for the technique of a grade-scale construction which may be summarized as follows : supposing an examiner desires to make a scale for addition for the third grade :—

(1) He selects according to his judgement a number of problems varying in difficulty.

(2) He tests the problems with a number of third-grade pupils chosen at random.

(3) He finds the percentage of pupils who solve each problem correctly, larger percentages obviously indicating less, and smaller percentages greater, difficulty.

(4) He tabulates the results and converts the percentages into P. E. units of difficulty.

(5) He calculates the P. E. distance of the zero point of addition ability from the third-grade median.

(6) He calculates how many units of P. E. each example is above the zero point, and his scale is complete.

(7) He sometimes chooses to delete from the scale problems which do not come at equal P. E. intervals.

(8) If his aim be the construction of a scale for the entire school instead of for one grade only, he repeats the second, third and fourth steps for each of the other grades.

(9) He then calculates the distance in P. E. units from each grade median to the adjoining grade median or medians. This is done by reckoning the percentage in one grade who score higher than the median of the adjoining grade. This percent shows the P. E. distance between the medians of the two grades.

(10) He then uses the intervals to compute the distance in P. E. units of each example from the common zero point of reference on the basis of its P. E. distance from its own grade median.

(11) He is then in a position to calculate the final elementary school P.E. value for each example, and thereby to locate it in the scale.

The age scale is another device for scaling the test the basis of which is the growth unit. In this case the desideratum is the attainment of satisfactory age norms. Supposing again we are wanting to construct a scale for measuring ability in addition, we first of all find what the average score for pupils of a certain age may be; or else the median score, if the median be the basis. Then we determine the performance of the pupil in question. If we find that his score is exactly that of the average or median for that of children of his own chronological age, then we say that his Educational Quotient, so far as ability in addition is concerned, is 100; if his score is 85 per cent of that of the average or median for his chronological age, we place his educational quotient at 85; if his performance is 115 per cent, we fix his educational quotient at that. The following table will illustrate how it may be set down in the case of a class which is being measured:

Test.	Age.								Pupil's Test Score	Pupil's Age.	Pupil's E. Q.
	8	9	10	11	12	13	14	15			
A—Average score .	4	8	12	15	18	20	22	24	15	11	100
B— Do. ..	4	8	12	15	18	20	22	24	18	10	150
C— Do. ...	4	8	12	15	18	20	22	24	15	13	75

A third method of scaling the test is the percentile scale. The percentile method, as the name implies, is an arrangement of scores of performance on the basis of percentages. We have already observed that this was the method which was adopted by Pintner and Paterson in their "Scale of Performance Tests." In speaking of it they have this to say: "The presentation of the results of tests in the form of percentile tables is a comparatively recent innovation in the history of mental tests. It has arisen naturally with the testing of large groups of individuals. The method would be impossible with few cases. It has arisen also, from a desire to know what the distribution of a group really is in respect to the various portions that go to make up the total group. Our belief

that individuals, in regard to all kind of abilities, distribute themselves on a normal curve with the very good ones at one end and the poor ones at the other, rather than into distinct types, is leading us to insist more and more upon a presentation of results that can be interpreted in this manner. The 25 and 75 percentiles so commonly used at present are the result of our desire to know what the middle 50 per cent of 'normal' group of the individuals tested can do. The addition of other percentile points gives us a finer means of discrimination. It has long been customary to consider the middle 50 per cent normal, the upper 20 or 15 per cent bright, the uppermost 10 or 5 per cent very bright, the lower 20 or 15 per cent poor, and the lowest 10 or 5 per cent very poor. The division into 10 percentiles will allow us to increase our groups greatly, and in time to attach a definite meaning to each of the ten percentile abilities."[1]

The method of constructing a percentile table is somewhat as follows. The scores of the various individuals who comprise the group are arranged in order of magnitude, and if the calculation be made in the direction of low to high, the 10 percentile is found by counting through one-tenth of the scores, the 20 percentile by counting through one-fifth of the scores, the 50 percentile by counting through one-half of the scores, and so on. While the percentile method does not serve as a criterion for fixing one's mental age or grade mentality, it enables us to make comparisons with the median of a group, and to learn how any individual stands with reference to the total group. There is one most obvious difficulty with the percentile method. It is customary to draw up percentile tables for each separate test. But that does not give a fair index to one's mentality, as his position in the various percentile tables may show great variability. If one is to reach any sound conclusion it is necessary to draw up over and above these percentile tables what Pintner calls a sort of "super-percentile" table which will indicate the true percentile value of the various median percentiles. It will be constructed like other percentile tables but the data will be the various median percentiles.

A fourth type of scale which has been devised whereby the tests may be scaled is the product scale. On this basis the performance of any test is scored as a product with reference to some samples or specimens which have been previously graded. This grading of the specimens may be either on the basis of the performances of adults or on the judgement of adults. We have had occasion in the chapter on Tests of Attainment to refer to scales of both types. Of the former type we noted the handwriting scale of Dr. Leonard P. Ayres. In fixing his criterion he parted company with Professor Thorndike whose creation was "general merit," and substituted "legibility," at the same time claiming that such

1 pp. 184, 185.

a change involved substitution of function for appearance as a criterion for judging handwriting. The method of scoring any individual performance is to move it along the scale until it has been ascertained which one of the specimens is the best index of the quality of the handwriting of the individual, the pupil being given a mark of 20, 30, 50, or whatever it may be in accordance with the value placed upon the specimen to which it approximates. It may be urged however that this method of scoring is scarcely so objective as it may seem at first sight. The scale of products which is accepted as a criterion is, to begin with, fixed on the basis of judgements as to what constitutes legibility, a matter on which unanimity would be difficult to obtain. And in the second place the judgement of the individual performance with reference to the scale is also more or less subjective. At the same time in such a subject as handwriting it is difficult to conceive of any way of avoiding an element of subjectivity, and after all criticized public opinion is not such a defective brand of subjectivity.

Another attempt at a product scale is one that is plainly based on the variability of judgement. We observed a type of this scale in the Hillegas' English Composition scale. We may summarize the points which McCall enumerates in describing the construction of a scale of this kind:—

1. Specimens of compositions are selected by the scale constructor, ranging in merit from zero to ninety.

2. He then requests a number of competent judges to arrange them in order of merit.

3. He calculates from the percentage of judges who make the various rankings a table of rankings.

4. He then subtracts 50 per cent from all the percentages so obtained.

5. He then determines the P.E. difference in merit between each specimen and each other.

6. He makes P.E. calculations also in many indirect ways (E.g., NA=TN-TA).

7. The mean of all possible direct and indirect calculations of the P.E. differences is reckoned as the true difference.

8. Specimens are then arranged in order of merit on the basis of these calculations.

9. Record is made of the number of judges who give a zero-mark to each specimen.

10. The median zero specimen is then determined.

11. The P.E. distance of each specimen above the zero specimen is considered its scale value.

12. The selection of specimens above the zero specimen is such that the distances between the different specimens will be approximately of equal P.E.

McCall very well says that "education is interested in many kinds of differences," so that the product scale has its use in giving us another way of making educational calculations. There are absolute differences in such subjects as arithmetic which can be calculated on an absolutely objective basis, but there are other differences that afford no such basis, for comparison. They depend entirely on judgement and the nearest approach to objectivity which we can obtain is to obtain the judgements of a number of men who are admittedly experts, and to standardize their judgements. That is the way in which we must have scales constructed for such subjects as handwriting, composition, and drawing where there is room for differences in judgement. The fact is that the product scales which are devised for these subjects are not tests at all, in the sense that we usually speak of tests. They are rather techniques to enable the instructor to standardize his method of scoring.

A fifth type of scale for sealing the test is the "T" scale of McCall which was constructed on the advice of Thorndike as a means for the measurement of reading. His description[1] of the method whereby the scale was constructed may be summarized as follows:—

1. Selections of reading material, both prose and poetry, of graduated difficulty, were made.

2. Questions were framed on the basis of the text whereby the subjects could respond with brief, scorable answers.

3. Several experts answered all the questions, and assisted in arranging them in order of difficulty.

4. The test and its accompanying instructions were mimeographed.

5. The test was then applied to a few hundred pupils in grades III to VIII, in order to give data for the study of distribution of scores.

6. The scoring of answers was as either right or wrong.

7. Some of the questions were deleted as unsatisfactory, on the basis of the preliminary test.

8. The results of the remaining questions were tabulated by each question for each pupil.

9. The total number of pupils answering correctly each question was calculated and divided by the number of pupils tested to obtain the percentage of correct answers.

10. On the basis of this calculation the Standard Deviation difficulty was reckoned.

11. The questions were then rearranged in order of the actual difficulty as disclosed by the preliminary test.

---

[1] Cf. McCall: How to Measure in Education, chapter X.

12. Any serious gap of difficulty which could not be filled in by shifting the positions of questions was overcome by combining two or more questions into one.

13. The materials thus finally rearranged were printed in booklet form with which was included instructions.

14. A final test of the scale was its administration to a group of schools which were fairly representative of all ages.

15. Once more the test was applied to all pupils from grades III to VIII and special attention was given to all pupils between the ages of 12'0 and 13'0 in whatever grades they might be found.

16. The answers to each question were scored as right or wrong in accordance with a definite plan. Giving partial credits was not found to be very satisfactory.

17. All correct answers, the worst answers accepted and the best answers rejected, were tabulated to afford a scoring key.

18. The tests books of the different pupils were taken as the basis of classification according to ages and grades.

19. The total number of questions answered correctly by twelve-year-old pupils was calculated.

20. The percentage of twelve-year-olds who exceeded no questions plus half of those who did no questions was calculated. Similarly was computed the percent of those exceeding one question plus half of those doing one question; and again with two questions, and so on.

21. These percentages were converted into S. D. values or scale scores, and the results tabulated.

22. A table was constructed which indicated the number of pupils of each age answering correctly a definite number of questions in the interest of building up age norms.

23. The total number of pupils for each age, the total scale score for each age, and the mean scale score for each age was calculated.

24. The mean scale score is faulty both on the lower and on the upper sides because of the limits set by the scale itself. The investigator was certain that the means for the lower ages were too high, and those for the upper ages too low. There are technical statistical methods whereby the defects could be corrected and the true means discovered, but McCall believed that inspection, guided by the mean and the true mean was accurate enough. Represented diagramatically, the mean scale scores are a crooked line, and the true mean a straight line. The truer mean would still advantageously be represented by a straight line but with a little more deviation from the true mean in the direction opposite from the mean scale score, in order to correct the defects on the lower and upper sides.

25. A table was also constructed to show the results according to grades and for sections of grades.

26. Special attention was given to test sixteen-year-olds as had been previously done with twelve-year-olds.

27. In the community tested 20 per cent of sixteen-year-olds were in high schools which was taken to mean that this 20 per cent was the brighter portion of the sixteen-year-olds of the community.

28. The number of correct answers for 35, 34 and 33 questions was determined for the sixteen-year-olds.

29. To get the percentage of correct answers the number was divided not by the 20 per cent who were tested but by the 100 per cent of children of that age in the community.

30. These percentages were converted into S. D. values.

31. On this basis the scale was extended on the upper side. An extension downward was not felt to be needed.

32. The scale was then published—both the tests and a leaflet of directions for applying and scoring the test.

Summing up his findings after the construction and thorough application of the T-scale, McCall says:

"Thus the T-scale method was developed not only to provide a more satisfactory reference point and unit of measurement, but also to provide a method of combining scoring units which yields a geniune scale score for each pupil, which combines units by the method of simple total, which preserves all the original test material, and which is simple enough to be used by non-statistically trained educators. All these objects were attained at one stroke by scaling the total score . . . Scaling the total number of questions correct or, when more than one point is given for each question, the total number of points made shows immediately the scale score corresponding to each total number of points, which in turn is secured by merely adding the points made on the different test elements[1]."

II.—CORRELATION.

The second great problem in which we may profit by the findings of the statisticians is the problem of correlation. We may describe correlation as used in this connection as a statistical measure of the degree of correspondence between various particular abilities or between a specific ability and general ability. The term is also used in reference to the degree of correspondence holding between the findings of different tests as measures of the same ability. It is a term appropriated from geometry and expressing the mathematical measurement of relationship. We have already used the term in connection with the amount of correspondence both between abilities and between tests and scales of tests. Without some such device it would be very difficult indeed to attain any sound conclusions in regard to the

[1] Op. cit., p. 305.

whole task of standardizing measurements of either intelligence or progress.

One of the earliest forms in which men became interested in the subject of correlation was in regard to the correlation between brain and intelligence. It is one phase of the perennial problem of the relationship between body and mind. Various solutions have been suggested, but we need not tarry over discussion of them at this juncture. It may suffice to say that the argument has been pretty well narrowed down to a controversy between psychophysical parallelism and interaction. One interesting bit of evidence which at first blush seems to support the hypothesis of parallelism is that the brain reaches its maximum weight about the same time that the intelligence attains its maturity. At the age of fifteen the brain has attained its full weight, and at the age of sixteen the mind has reached its maximum development. Ballard gives an interesting proof of the latter fact which came out in his standardization of his absurdity test. When he reached the year sixteen the median or norm was a performance of 18·9 in a test of 34 parts, and thereafter the performance remained constant. But we must not be too hasty in concluding that these facts justify the theory of parallelism, because these are not all the facts. On the same principle the brain of a child of five should be much smaller than that of an adult, but on the contrary it is 90 per cent of its maximum size, and the brain of a feeble-minded person ought to be much smaller than that of a genius which it is not. It is therefore impossible to establish any significant correlation between the weight of the brain and the amount of intelligence.

Correlation is a statistical measure of the degree of correspondence, whether between general intelligence and specific abilities, or between intelligence and school achievements, or between two sets of mental tests. Our interest in correlation is therefore from at least three angles of approach. In educational matters there is much that can be done in the measurement of relationships without requiring the use of such exact measures as the method of correlation. One simple method is that of plotting by which distributions may be diagramatically represented, and the relationship between two series shown in the form of a graph.

The problems mentioned call, however, for a more exact type of measurement than can be secured by means of a graph. Correlation is a statistical instrument that affords that exactness. It gives us not only the relation of one quantity, say A, to another, say B, nor only the relation of B to A, but a peculiar composite of both of these relationships taken together. To quote Thorndike: "A correlation is a mutual, not a one-direction relation; is not the relation of absolute amounts of divergence, but is the relation

---

[1] Mental and Social Measurements, p. 160.

of such amounts divided by the variability of the trait in question; and assumes, in so far as a single co-efficient is to be its adequate measure, that the relation lines for A to B and B to A are rectilinear."[1]

Various statistical formulae have been devised whereby the correlation between two factors is measured. I shall mention two of the more commonly used ones, and illustrate them from a simple case. The one is known as the Pearson formula, thus,

$$r = \frac{\text{Sum of } (x\ y)}{N\ \sigma_1\ \sigma_2}$$

where x and y stand for the deviations of each of the measures from the mean value of the series, σ_1 for the standard deviation of the first series, σ_2 for the standard deviation of the second series, and N for the number of things or persons measured.

The second formula is given as

$$R = 1 - \frac{6 \times \text{sum of } D^2}{N\ (n^2 - 1)}$$

where D denotes the difference between the two integers which indicate the position of the two related measures in their respective series, and n denotes the number of pairs of related measures.

Let us suppose, for example, that we desire to know the reliability of the Stone Reasoning test in arithmetical ability. We administer it on two occasions to the same set of nine boys, at two periods, one year apart. The boys have all had the same opportunity to make progress in arithmetic during the intervening year. If the test were perfectly reliable, then they ought to make progress at a rate sufficiently equitable to secure a fair degree of positive correlation between the two applications of the test. If the result of the two applications of the test was that the boys stood in exactly the same order of rank on the two occasions, then the correlation would be perfect or + 1; on the other hand, if the order were exactly reversed in the second performance it would mean that the correlation was inverse or — 1; if the data with which we had to deal were of such a nature that we were unable to reach any conclusions at all, we would describe the correlation as zero. Any amount of positive correlation, be it never so small, indicates some correspondence, and the greater amount of positive correlation the greater the amount of correspondence is thereby indicated, until we reach + 1 which indicates a perfect correlation. Conversely any amount of negative correlation indicates that the correspondence is in the direction of inverse relationship, until we attain — 1 which denotes an exact inversion of the two series under comparison. Concerning these fundamental facts, all of the formulae are agreed.

But let us proceed with our hypothetical case, in which case we shall suppose a real problem which we cannot answer by merely

observing the data. In cases either of perfect positive or perfect negative correlation, we would obviously not need any mathematical formulation to help us to reach our conclusion. We are then going to measure the reliability of the Stone Reasoning test by its application to a class of nine boys on two occasions. Let us suppose that our results were as follows :—

Individuals tested.	Rank of each individual in first test.	Rank of each individual in second test.	D	D^2
Ramaswamy	3	1	2	4
Gopal	7	4	3	9
Krishnan	5	2	3	9
Abdul	8	6	2	4
Ratnam	2	5	3	9
Venkatayya	1	3	2	4
Ranganathan ...	9	6	3	9
Govindan	4	7	3	9
Subbiah	6	8	2	4

$N = 9$ sum of $D^2 = 61$

$N^3 - 1 = 80$ $6 \times$ sum of $D^2 = 366$

$$\text{Correlation is } 1 - \frac{6 \times \text{sum of } D^2}{N(N^2 - 1)}$$

$$= 1 - \frac{366}{9 \times 80}$$

$$= 1 - .508$$

$$= .492$$

Calculating the same problem on the other formula,

$$\text{Correlation is } \frac{\text{Sum of } (x\,y)}{N \times \sigma_1 \sigma_2}$$

$$= \frac{-2x4 + 2x - 1 + 3x1 \quad -4x2 + 4x1 + 1x2 + 1x3}{9 \times \sqrt{6.67} \times \sqrt{6.67}}$$

$$= \frac{22}{60}$$

$$= .37$$

In the preceding section some attention has been given to the characteristics of a test. It ought to be apparent that the manner

of examining into all three of these characteristics — validity, reliability, and objectivity — is by the use of these statistical formulae. The best way of discovering whether a test really measures what it purports to measure, and measures that factor consistently, is to determine the co-efficient of correlation between various tests of the same ability. Take the example of the completion test in which certain words are omitted from a passage, and the subject is required to fill in the omissions with words that make sense. It is quite apparent that this performance calls into function the association processes. But there are several tests, such as the analogies test, the test of completing pictures from which features are missing, the rhyming test, etc., which call into function the same processes. If we wish to test the validity of a completion test that we have devised, one way to do it is to determine the co-efficient of correlation between our test and other tests such as those indicated which call into play the same ability. Or again, it may be done by determining the co-efficient of correlation between our test and another test devised by some other investigator which is one of the same ability. The same method may be applied to the testing of the validity of a scale of tests. I have alluded to the fact that the Army psychologists found the proof for the validity of their newly devised group tests by establishing the fact of their high positive correlation with other scales in existence such as the Stanford-Binet and the Point-Scale.

Again as to the matter of reliability, we find the same method standing us in good service. The amount of agreement that subsists between results obtained from two or more applications of the same test to the same pupils by the same examiner can only be determined with precision by means of the co-efficient of correlation. In the hypothetical instance which I have given, this is the type of case that is illustrated. A really reliable test, as I have stated before, should have a correlation between its various applications approximating + 1.

Furthermore the one way in which to determine whether a test is objective, or whether it is purely subjective, is to have it administered by different examiners to the same subjects, and then to calculate the co-efficient of correlation between the results of the trials. If the correlation be positive and high, we have the best possible evidence of the test's objectivity ; if it be zero or low, the evidence points in the reverse direction. So that in the case of all three tests which we wish to administer to the tests themselves, the instrument for precision is the method of correlation. At the same time it ought to be observed that in the hypothetical case taken, the number of those supposed to be tested was insufficient for adequate results. The number ought to be at least 15 or 20, and preferably more. Not only so, but the individuals should be

representative of the population or group tested, and the tests should be so selected as to call forth an adequate range of abilities. Nothing but the greatest care can be expected to yield results which have the character of mathematical precision.

Nobody claims infallibility for the tests, even when the rigidest mathematical processes are employed in working out the degrees of correspondence. But we are able to ascertain with accuracy the limits within which the probability of error will fall. In that way we can demonstrate that the psychological test of mental ability has greater value in diagnosis and in prediction than any other instrument yet devised. It will be noted that the "median" has been spoken of much more frequently than the "average." It has been ascertained through experience that the "median" is more satisfactory because it is less affected by eccentric performances that are likely to occur at either end of the line in mental testing. It is therefore more representative of the whole population. The average is more easily computed in many cases, as it is found by dividing the total scores by the number of performers. But the median is computed by arranging the scores of all the individuals in order of merit, and then counting off from either end until the middle individual is reached; his score is the median. The probability is that, if the number tested were very large, there would be no great difference between the median and the average, but if there were a marked difference it would probably be due to some unusual performances either by geniuses or by blockheads or by both, throwing the average away from the middle.

One positive result may be noted as an outcome of studies in correlation. The evidence goes to show that there is no antagonism between various types of ability. On the other hand there is a good deal of evidence to indicate that many specific abilities have little or nothing in common. For example, mechanical skill and general intelligence, though they show positive correlation, do not show a high correspondence, the measure being about .4. Wyatt investigated the amount of correlation between the ability to interpret fables in the Binet scale and the ability to put together dissected pictures, and obtained a result of .26. On the other hand the correspondence between the test of association by cause and effect and that of general intelligence yields a positive correlation of from .85 to .94 — an almost perfect correlation. The opposites test yields a correlation of .96 with the tests of general intelligence. Whipple's *Manual of Mental and Physical Tests* gives a great deal of data in regard to correlations which have been worked out between different specific tests, and between tests of specific abilities and general intelligence.

CHAPTER X.

PRACTICAL PROBLEMS FOR THE INDIAN EDUCATOR.

A study of conditions in India will make it evident that the same types of needs exist here which led to the introduction of the science of mental measurement in France, England, the United States, and other countries. There is the need which is created by the lack of accuracy in regard to school marks and in regard to examinations. There is further the need brought about by the lack of any scientific method of classifying mentality both in schools and elsewhere. There is the ever-present problem of retardation. And then there is the complex problem connected with the mental phase of the problems of crime, delinquency and disease. In all of these situations the need for greater accuracy in calculating mentality quantitatively is being felt. So that to all of these situations our science applies with peculiar cogency.

The first problem is that of inaccuracy in the system of marking, the injustice of which is most obvious in examinations. This is a difficulty which the inspecting officers are continually encountering as they visit the schools of this Presidency. To be in the fourth standard means quite a different thing in one village from another, for the standards of marking and of promotion are very different. We have here one of the reasons why some schools show up well in the public examinations and others poorly. If standards have been kept too low and the fear of losing popularity or of offending fond parents has led to too easy promotions, which is frequently the case, the evil of the policy may lay quite dormant until a public examination comes and the school makes a disgraceful showing. I think of a High School that I visited where the Headmaster had been too generous in his promotions from year to year until more than half of the Sixth form was composed of pupils who were not prepared for it. Only ten per cent passed in the School Final examination, I believe largely because of the lack of standardized measurements being used in the school organization. The probability is that our Inspectors could repeat to us many examples parallel to this one.

Even where an effort is made to maintain a recognized standard, if there be no adequate unit of measurement, there will have to be a generous allowance made for differences in interpretation.

Our friends, Professors Seshu Ayyar and Ranganathan of the Presidency College, have been investigating examination results here in this Presidency. In the introductory paragraph of their article in the *Journal of the Indian Mathematical Society* on *A Statistical Study of some Examination Marks* (April, 1922), these investigators say :—"It is a well-known fact that, with all the care that is bestowed upon it, the standard of the question paper is not the same from year to year, neither can the valuation be regarded as standardized. It is therefore very desirable that some method be found to make due allowance for these unavoidable variations (in standards) so that candidates may not suffer and the value of the examination as a test of fitness may remain steady. " Further, from the pedagogical sta dpoint, it would be of interest to get some quantitative measures of the correlation of the candidates in the various subjects."

With this in view these gentlemen investigated the results of a certain public examination for six successive years in certain subjects, and subjected their findings to statistical treatment. Their investigation dealt in some detail with the minimum required for a pass which led them into a discussion of the margin of P. E. " Justice requires that candidates whose marks are lower than the adjusted minimum by less than the probable error must be given the benefit of the doubt." In calculating what allowance should be made for probable error, they propose to adopt the findings of Professor Edgeworth because of the absence of any such calculations for Indian conditions. Professor Edgeworth's classifications of the causes of probable error, and the distribution in accordance therewith is as follows :—

(i) *minimum sensible* which is defined as " error due to the difference of perception of excellence whose magnitude varies with the subject, being least in Mathematics and perhaps greatest in Composition " and in this instance reckoned at 7 per cent in Physics and Chemistry and 10 per cent in English and History ;

(ii) *personal equation* which is calculated at the rate of 10 per cent on the mark of each answer ;

(iii) *difference in the scale adopted by the several assistant examiners* which is here computed at 4·5 per cent for each paper ;

(iv) *fatigue of the examiner* for which an allowance of 1·5 per cent is made on each paper ; and

(v) *speed of valuation* which is computed at 25 per cent on the mark of each paper.

Taking all of these matters into consideration in connection with their particular investigation, Professors Seshu Ayyar and

Ranganathan have concluded that the aggregate probable error may be taken as the following percentages of marks earned by border-line candidates :—

English	Mathematics	Physics	Chemistry	History
4·4	4·8	5·1	5·1	5·8

There is nothing at all surprising in the results of this investigation. Indeed one would be tempted to prophecy that a still more extended investigation into the results for the various examinations covering a longer period would disclose a greater degree of variability. It is particularly surprising to find English giving a smaller percentage of probable error than Mathematics. It is what might be anticipated when we find these men endorsing the recommendation of the Calcutta University Commission for the appointment of a skilled statistician to the Board of Examinations to help to overcome this difficulty.

But the trouble is more deep-seated than in the marking of examination papers. It is there because it is elsewhere. It is due to the lack of standardized examinations, a trouble that exists all along the line from the lower elementary grades to the higher University examinations. Under the prevailing circumstances the variations which exist are indeed "unavoidable," but if there could be devised a complete series of attainment tests, thoroughly standardized and tested for validity, reliability, and objectivity, a large proportion of the present variability could be overcome.

The vexation of retardation is with us as well as with educationalists in the West. Certain investigators in the United States have collected statistics which lead to the conclusion that about 25 per cent of school children in that country are retarded. It would be most useful if somebody would take the matter up for investigation in this Presidency. Terman calculates that a sum equivalent to more than a crore of rupees is expended annually by the United States for the re-education of backward children. How much is the Madras Presidency expending annually for the same purpose? Whatever the amount may be, certainly a fair proportion of it might be saved, if we had standardized mental measurements which would give us the evidence that we want as to whether a pupil is mentally capable of going on any further than he has gone already, or whether he has reached the limit of progress as far as school is concerned. There are boys being kept on year after year in some of our High Schools without promotion, or, if with promotion, it is because they are pushed on rather than because they have succeeded in the tests, and whose continuance in school is either because the school authorities want the fees, or the boy is a good hockey player, or because his parents are persons of influence in the community, or for some other such fatuous reason. If psychological testing were done regularly and

the results put to practical use, such anomalous situations would largely disappear.

The connection between crime, delinquency, and disease on the one hand and mental defectiveness on the other hand is a broad field for investigation. Intelligence is largely a matter of native equipment. It depends upon neurological factors which may in turn depend upon physical and chemical processes in the nervous system, particularly in the cerebral cortex. But heredity plays such an important part in the determination of one's intelligence that for eugenic reasons the State ought to take a more lively interest in this problem. It has been disclosed that insanity, imbecility, mathematical genius, musical ability, and so on, are frequently family characteristics. It stands to reason that if certain characteristics are dominant on both sides of one's progenitors, such characteristics should continue to be dominant in their progeny. The investigations of the inmates of jails, homes for delinquents, hospitals for alcoholics, houses of ill-fame, etc. disclose the fact that there is a high degree of correlation between mental defects and moral defects. Examinations of children, especially where there is compulsory education, will bring to light cases of feeble-mindedness and enable the State to deal effectively with the matter. It will also contribute to a more scientific classification of mental disorders. Here in Madras the term "insanity" in official language seems to be the all-inclusive term for every mental defect—rather a sad comment upon our modernity.

I.—The Problem of Training.

One of the first difficulties that we feel here in South India in getting on with work in mental measurement is the lack of men who possess the necessary technique. In spite of the fact that some psychologists are less enthusiastic about the validity of the tests than others, still there is general agreement that, with all their defects, they afford a better criterion for measuring the human mind than any other device that has yet been constructed. The question arises, however, as to what extent specialized training is necessary for the application of psychological tests. Concerning that matter there is no unanimity. Some claim that a very thorough training is required, for if the testing be put into the hands of inexperienced persons, no matter how enthusiastic they may be, the results will be of doubtful value. Others claim that they have so constructed their tests that the most inexperienced teachers may, by simply following the directions, achieve perfectly satisfactory results. Still others tend to a middle ground, saying that the experimenter ought to have training, but that a six weeks course with competent instructors and plenty of object-lessons would suffice to prepare a person for independent work.

The difference of opinion on the subject of training is not one which will disappear merely on the basis of any amount of argumentation. The only way to come to any conclusion is the way in which we had to reach conclusions in regard to the tests themselves, *i.e.*, by experimenting. A comparison of the results achieved by untrained or meagerly trained examiners with those obtained by men of thorough training would be the only sure way of reaching valid conclusions as to the necessity or otherwise of thorough training. One investigation of this kind was made and reported in *The Training School Bulletin* for 1914 (pp. 113—117), by Dr. Samuel C. Kohs.

"Dr. Kohs gives the results of tests made by 58 inexperienced teachers who were taking a summer course in the Training School at Vineland. The class met three times a week for instruction in the use of the Binet scale. During the first week the students listened to three lectures by Dr. Goddard. The second week was given over to demonstration testing. Each student saw four children tested, and attended two discussion periods of an hour each. During the third, fourth, and fifth weeks each student tested one child per week, and observed the testing of two others. The student was allowed to carry the test through in his own way, but received criticism after it was finished. Twice a week Dr. Goddard spent an hour with the class, discussing experimental procedure. The subjects tested were feeble-minded children whose exact mental ages were already known, and for this reason it was possible to check up the accuracy of each student's work.

"Kohs' table of results for the trial testing of the 174 children showed:—

(1) that 50 per cent of the work was as exact as any one in the laboratory could make it;
(2) that in an additional 38 per cent the results were within three-fifths of a year of being exact;
(3) that nearly 90 per cent of the work of the summer students was sufficiently accurate for all practical purposes;
(4) that the record improved during the brief training so that during the third week only one test missed the real mental age by as much as a year.

"Since hardly any of these students had had any previous experience with the Binet tests, Dr. Kohs seems to be entirely justified in his conclusion that it is possible, within the brief period of six weeks, to teach people to use the tests with a reasonable degree of accuracy.

"What shall we say of the teacher or of the physician who has not even had this amount of instruction? The writer's experience

forces him to agree with Binet and with Dr. Goddard, that any one with intelligence enough to be a teacher, and who is willing to devote conscientious study to the mastery of the technique, can use the scale accurately enough to get a better idea of the child's mental endowment than he could possibly get in any other way. It is necessary, however, for the untrained person to recognize his own lack of experience, and in no case would it be justifiable to base important action or scientific conclusions upon the results of the inexpert examiner. As Binet himself repeatedly insisted, the method is not absolutely mechanical, and cannot be made so by elaboration of instructions."[1]

The consensus of opinion seems to be that even untrained examiners, who will devote themselves to a careful study of the tests, can secure results which will give them a surer index to a subject's mentality than they can secure by any other means. And further, that within a short course, say six weeks for a graduate, it is possible so to master the technique as to be able to secure as valid results as any one else. But it should be added that further training must not be construed as waste of time. The better the examiner is trained, the less mechanical will be his procedure, and the more will be his ability in interpreting results. Although much information may be gained by those who are not well trained, still all the training which it is possible to secure will be found useful. One cannot be too close a student of psychological processes in work of this nature. Moreover, the two phases of the theoretical and the practical will be found to work together to the mutual benefit of both. All the knowledge which we possess of psychological processes and functions will be found to make the work of mental testing much more significant to the examiner. And conversely, all that an examiner may be able to do in actual testing of subjects as to their mental abilities will be found useful in unfolding the working of the processes. We begin our measuring with a tentative definition of intelligence, for example, but when we conclude we have information which will put us in a much better position to achieve a satisfactory definition.

Here in South India the problem is acute because, though there is a keen desire to get along with some work in this direction, there are very few who know enough about it even to make a beginning. The purpose of this course of lectures, I take it, has not been to try to throw new light on the problems of mental measurement with which men have been struggling in other countries, but rather to give a little information whereby interest in the subject may be awakened, and those who are interested may know something about how to proceed in a tangible way.

---

[1] Terman's book on *The Measurement of Intelligence* gives a summary of the report (pp. 107—109). I have quoted from Terman's account.

Fortunately there a few persons scattered throughout India who have had experience and training in this field in Western Colleges, and who are bringing their experience to bear upon our problems here. For the present perhaps we shall have to depend upon these individuals to begin the work and to point the way to others. I have made some references here and there to some efforts that are being made. I referred, e.g., to the experiments of Rev. D. S. Herrick of Bangalore with the Goddard Form-Board. In the Narsinghpur High School, Central Provinces some work has also been done with the Form-Board test, as also in the College and Schools of the American Arcot Mission at Vellore. Some of these same workers have been experimenting with the Cube test also, and are gradually gathering data for the construction of norms. Quite a number of scattered workers have been experimenting with the use and adaptation of the Stanford-Binet tests. In the Government Training Colleges both in Madras and in the Central Provinces something has been done, while individual workers have been at work in various parts of this country, including Burma. Experiments have been conducted in some centers with the Achievement Tests, as e.g., with the Ayres' Spelling Scale, the Courtis Arithmetic Scale, the Kansas Silent Reading Tests, etc., but these efforts have been even more scattered that those concerned with measuring intelligence. At present a movement is on foot to secure some sort of clearing house arrangement, so that the results of all that is being done may be collected, and that norms may be built on the basis of results that are as far reaching as possible, and further so that unnecessary duplication of effort may be avoided.

But something still more effective needs to be started if progrees is to be made in keeping with the demands of the present situation. If it were possible for the Department of Education to appoint some person who is a specialist in this field to give courses of lectures with experiments at the Training Colleges, and to travel to some extent throughout the Presidency getting work begun and organized in various centers, it would be well. If Dr. Goddard was able to give courses covering six weeks at the end of which time the students were able to work independently, why should we not have a number of special sessions covering the same period at different centers throughout this Presidency for the training of teachers, and possibly also of medical officers in this work? Perhaps the Government will tell us that they cannot afford it. They ought to realize that they can ill afford not to do it.

Certainly an adequate course in Mental Measurement ought to form part of the curriculum in every Teachers' College. Our teachers ought to be trained in the administration both of group and individual tests, both of language and performance tests, both in intelligence and in attainment tests. It is the best and indeed the only adequate method whereby we can look forward to standardizing

our examinations. If we want to standardize our examinations, certainly we must train our teachers in such a way that they will be in possession of the technique of measuring mental abilities. Furthermore we have observed that there is no such instrument for the detection of errors in teaching method or in student comprehension. Both for the purposes of diagnosis and of measuring the results of teaching, there has never been devised any method comparable to the methods of mental measurement. Professor John Adams in his recent book, *Modern Developments in Educational Practice*, has pointed out that one of the results of the knowledge which we gain in this way is the ringing of the knell of class-teaching. Too often in the past the class has been considered as the unit of instruction, the pivot around which the whole educational system has been made to revolve. But the tests have made it clear that there are individual differences which are too great to be neglected in this way, and not only so but that there are differences within individuals themselves which cannot be neglected in scientific teaching. Since abilities are plural and special, there is no adequate reason why a child should be compelled to take the work of a single grade in all subjects. In the more progressive institutions in the West provisions are being made to allow a child to make normal progress in all subjects. Perhaps that will mean taking arithmetic with the fourth grade, and reading with the eighth grade, or it may involve differences even wider than that. What of it? Education must have the child at its heart, and if it be not for the child, it has no right to be carried on in its existing forms. It may cost the State more to educate along these more scientific lines, but surely it is the wisest investment that a State ever made to spend its resources on its future citizens. To refuse to do that is to mortgage its own future.

II.—The Problem of the Tests.

A second great problem that faces us here is that of the types of tests which we shall find it the best to use. Are the tests that have been devised in the West suitable for use here, or shall we need to adapt them to Indian conditions, or must we construct entirely new tests? This is the problem of the test.

It will be evident to anyone who thinks for a moment that one of the difficulties that was found with the Binet tests is especially active here, viz., the language difficulty. There are two reasons why the language difficulty is a real one: (i) because there are so many illiterates, many of whom we shall want to test when the work is well started, as, e.g., criminals and other delinquents; and (ii) because there are so many different vernaculars that there is no one language which can be used as a medium of testing. The Binet tests are in French, and most of the revisions are in English, though some are in German. For the lower standards and for all

illiterate in English in India these tests are obviously defective. Something has been done to adapt them to the needs of the Indian situation by adaptations in Tamil, Telugu and Hindi, and perhaps of other vernaculars of which I have not heard. A real effort is being made to adapt and not merely to translate the tests. A translation would obviously be very inadequate, because the Binet tests call for responses to situations that are quite foreign to an Indian child. Let us take, e.g., such a test as the naming of words which rhyme with the words *day*, *mill*, and *spring*, a test included in the nine-year-old tests of the Stanford revision. To test the same ability, what is needed is the selection of three words with which it would be equally difficult and equally easy to rhyme words in Tamil or Telugu as the English words, for it would be quite another test to call for rhyming words to go with the translations of these words. Again the syllable-repeating test would plainly have to be one in which the child is given a sentence in his own vernacular, and not in a foreign language. But even with all of these precautions as to adaptation, it is very difficult to know whether or not the test calls for the same degree of mentality as the test in the other language. It is quite possible that the test may call for either an easier or a more difficult response in one country than in the other. Nothing short of a prodigious amount of experimenting followed by careful statistical treatment will enable us to form any adequate judgement on the merits of an adapted test as compared with the original. To be sure, it may be a valid test of mentality, and have also the characteristics of reliability and objectivity, so that it may find a place in a scale which we use here. What I am saying is that it would be unsafe to compare it on the level with the test of which it is an adaptation without putting it to a searching examination. After all we have no right to be talking about the testing of nine-year mentality in the United States and England and India and China, unless we have found from experimenting that the tests which we are using are testing the same level of mentality in regard to a particular process.

Another difficulty with the Terman test in adaptation to India is created by our immediate need. The Terman revision of the Binet test is for individual work, and is therefore a bit cumbersome, and requires a good deal of time to measure any large number of subjects. If we had to go through a school with say one thousand students, even though forty minutes to an hour is all that is required to measure each individual, it is plain that it would require a great deal of time. After we have succeeded in training some hundreds of our Licentiates in Teaching to do the work, such a task will not assume such magnitude. But at the present stage, it seems to be more desirable to use that new instrument for testing which has been devised since the Binet tests—the group test. In this way we shall be able to sweep through the schools at a much more rapid rate

and thus gain a rough idea of the mental status of large numbers of children, especially in the elementary grades where we require the information to enable us to direct their work along wise lines. In this way we can make use of the group test for our first preliminary survey, and utilize the individual test for particular testing of individuals concerning whom we require more detailed information. The group test will enable us to discover those of low mentality, and thereby give us just such information which we need to save our time and money in trying to pull along those who cannot be led because of inherent incapacity. For all of these reasons, we need such an instrument as the group test that will enable us to do our preliminary work on a wide scope.

But that is not the only difficulty with the Terman individual test. It has also the language difficulty to obviate which the performance tests were devised. The question which may well demand some of our attention is that whether or not a scale of performance tests would not be more suitable to conditions here. Some are quite convinced that such is the case. We want to test all communities, all degrees of literacy, all ages, all language areas, and so forth. No test that depends for its application on literacy in any language would suffice. Nor would a test such as the Pintner and Paterson Performance Scale, where individuals have to be tested one by one suffice to give us a mass of information within a short period. It seems to me that the type of test best adapted for immediate needs is a group test of the performance type, something akin to the Army Beta Scale. This would combine the opportunity of collecting a large amount of data within a short period with that of obviating the language difficulty, and at the same time would be possible of application without regard to environmental differences.

The question of environmental differences is one of the factors which must determine the selection of tests. A test which would allow a Brahman child to score 100 per cent, and on which a Panchama child would get zero would obviously not be measuring intelligence, but would be simply accentuating the difference in social opportunity. Or a test on which a literate boy would score high and an illiterate child could do little or nothing would plainly be measuring not mentality but schooling. If we are to measure an ability, and to make legitimate comparisons on the bases of our measurement, the only fair criterion must be one which gives an equal opportunity to the poor and the rich, the high caste and the non-caste, the literate and the illiterate. Mr. Herrick's experiment to which reference[1] has been made heretofore, seems to point the way to the type of test which may very well meet the needs of this situation. The performance of the response required by

[1] Pp. 93, 94.

the Goddard Form-Board needed no specialized kind of knowledge which would be obtained through schooling or other experience It did not even require the use of language, for, although a few words of instructions were given in this experiment, still it would be quite possible to make the correct response with no other instructions than signs or gestures to proceed. The experiment showed that the average time for a Panchama child was two and one-half seconds longer than the average of a Brahman child in an experiment for which five minutes was allowed. So that from the point of view of the environment this experiment confirms the hypothesis that the best type of scale with which to begin work in India will be a performance scale, similar in type to the Beta scale used in the American army.

In considering the selection of tests, we have to bear in mind the particular character of abilities which we are measuring. Some are mental; others are rather motor. Some are of the type that involve the mental manipulation of the data on the basis of which the response is made; others are more practical and responses are made through a mechanical manipulation of material. Custom and tradition has in the past played a large part in determining the occupations of Indians. The influence of caste environments has been large. But we ought not to conclude *a priori* that individuals cannot do anything other than what tradition would assign to them. There is no way of reaching valid conclusions except by actual experiments as to whether or not, e.g., Brahmans possess motor abilities as well as mental, and Sudras mental abilities as well as motor. And the tests will have to be selected along lines broad enough to enable us to make comparisons of all the types of ability. We must get away from the traditional fashion of saying that a certain community is more capable than another. Capability is not a general abstraction which can be so lightly compared. If we are informed that a certain class is possessed of more ability than another, our response ought to be, "Able for what?" If abilities are specialized, as the experiments seem to indicate, then each one must be measured, and allowed to stand on its own merits. There is a tremendous amount of information which we ought to have before we begin to make general remarks about the various abilities, and about those in whom they are dominant.

The selection of the tests for use in India involves also a great deal of labour in regard to tests which shall measure achievement and progress. An attempt has been made to devise scales to measure handwriting, silent reading, and vocabulary in Hindi. These and other abilities, such as composition and spelling, need scales wherewith to measure abilities in all of the Indian vernaculars. When we see how much labour has been expended in the

United States to work out some of the existing scales, we must be prepared to do some work that will demand much of time and patience before we can build up satisfactory norms. But there is no other way to accomplish our end, and when we recall that it is the most precious product with which we are dealing—the child—surely the expenditure of time and energy is well worth what we give thereto.

The selection of vocational tests is another phase of the general problem of selecting or constructing tests. There may be some tests which have been devised in other countries which will meet certain situations here, especially where we are dealing with fitness for the same vocations. For example, it might easily be found that Münsterberg's test for the selection of candidates for telephone operators might be adapted, if not adopted, for use here. But there are other vocations for which tests would be of great usefulness, if they were scientifically constructed. Take, e.g., telegraph operators, type-setters, machine hands, mill hands, tram-car motormen, and other vocations calling for abilities both motor and mental, — efficiency might be greatly increased and money saved in the attempt to train impossible candidates, if there were tests used for the selection of candidates.

In all of these matters, one of the paramount points is to secure as much co-operation as possible among all who are working in this field. I need scarcely repeat that it is only by the collocation of an immense amount of data that satisfactory norms can be constructed, and that the tests can be validated. It is quite possible that in an area so large as India, with so many geographical and language divisions, a great deal of unnecessary duplication might take place through lack of co-ordination among those working in the various phases and problems. It is therefore eminently desirable that a clearing house should be established somewhere, and that efforts be made to secure information from all parts of the country where experiments of any kind are being conducted.

At present we have no adequate way of making comparisons between the school work being done in various parts of our own Presidency, let alone throughout India. True we have Government syllabi which the various schools are expected to follow. But even these are Provincial. Supposing we had the best correlation between schools in Madura and Bellary and Cocanada, we would then have no data for comparing a certain standard, say the seventh, in Madras and in Bombay or Lucknow. Even supposing the syllabi remained different for the various provinces of the country, it is immaterial so long as we could have standardized units of measurement whereby we could compare progress and achievement and abilities of different sorts. We have a general idea that standards in

Madras rank high when taking into consideration standards in the different parts of the country. But we have no adequate means of knowing exactly to what extent our general impressions are correct. The probability is that an analysis such as we could make on the basis of mental measurement would disclose that we were strong in some particulars and weak in others. Such an analysis would be much more valuable, because of its diagnostic significance. It would enable us to introduce such corrective measures as we found necessary into the system of instruction to bring our school system up to a high level in all branches.

III.—The Problem of Intelligence.

Intelligence tests are based upon the principle of sampling. In the same way that our friends of the Agricultural College at Coimbatore can determine the productive value of a piece of farm-land by a few samples of its products under determined conditions, so the psychologist claims to be able to form a reliable appraisal of one's intelligence by a few sample performances which call into function specific processes. The broader the range of the samples, the more conclusive will be the findings of the psychologist as well as of the agricultural expert. Standardized tests of intelligence therefore include tests of association, visual and auditory perception, time and space orientation, memory, comprehension of language, eye-hand co-ordinations, arithmetical reasoning, ingenuity, speed, ability to form concepts, and so on.

The present problem before us is to take samples that will enable us to judge of the functioning of mental processes among Indian subjects. Group psychology has made it apparent that there are certain differences between the peoples of different races. Mental measurement is going to enable us to judge to what extent these differences are due to environmental factors and to what extent the causes are congenital. Not only that; it will enable us to ascertain with much more precision what the factors are that differentiate the various group minds. In particular we are interested in finding out what are the component factors, or rather what are the dominant characteristics in the Indian consciousness. So far I am not aware of many investigations that have been carried out to determine these factors scientifically. But I may mention one or two investigations on the basis of which it is possible to make a few conclusions.

Professor John S. Hoyland of Hislop College, Nagpur, Central Provinces, made an investigation [1] in regard to the characteristics of

[1] Published by the Christian Mission Press, Jubbulpore, 1921.

Concerning the development of the Indian adolescent mind, he found that:—

(1) at ten years cruelty and fear are more noticeable than at other times;

(2) at eleven, the tendencies to save money and to self-interest are strong;

(3) at twelve, the mind of the child is more materialistic in ambitions and motives, and more egoistic than at other times;

(4) at thirteen, intellectual, ethical and religious interests come to the fore;

(5) at fourteen, conscience and intellectual interests are strong;

(6) at fifteen, hero-worship, conscience, altruism and the religious attitude are strong;

(7) at sixteen, altruism and religious motives are at their maximum, and bravery and loyalty are becoming stronger;

(8) at seventeen, intellectual interests and the aesthetic tendencies are strongest, while disregard for law and discipline are also at their greatest;

(9) the ages from 13 to 16 inclusive are the critical years in the adolescent's development.

Mr. Hoyland also makes certain comparisons which show variations with the dominant characteristics which similar experiments have disclosed in the West. His points include the following:—

(1) the Indian child is more susceptible than the Western child to the influences of morals and religion;

(2) the ethical ideals and the ambitions of the Indian adolescent are more vague than those of the Western child;

(3) Indian children have less idea of the meaning of public spirit, and less of the love of truth for its own sake, though their motives for lying are more altruistic than in the West;

(4) the Western child is more interested in animals than the Indian;

(5) the beauty of Nature appeals less and the beauty of architecture more to the Indian than the Western child;

(6) the critical faculty is more developed in the West;

(7) home discipline is weaker and school discipline stronger in influence over the Indian than the Western child;

(8) Indian children are more docile than Western;

(9) Indian children are more improvident with money than Western;

(10) altruistic considerations and the desire to obtain a good education are more appealing to the Indian adolescent than to the Western.

The tables and conclusions of Mr. Hoyland were based on 1,164 answers to Test A, 305 answers to Test B, and 436 answers to Test C. This involves a good deal of work, and yet it is not a very large bulk of material on which to generalize in any more than a tentative way. The value of the investigation is therefore in the preliminary conclusions, and should be carried on further in various parts of the country with larger numbers, before we can generalize on a wider scale.

During the academic year 1921—1922 the present writer carried on a number of investigations in the association processes, among the students of two or three colleges in the city of Madras. In particular the experiments were in uncontrolled or free association, though a few experiments were also conducted in controlled association. The purpose was to ascertain experimentally what were the dominant associations in the mind of the Indian student. Tables of results are available in Whipple's *Manual of Mental and Physical Tests* on the basis of which it was possible to make comparisons with results which were obtained by investigators who had performed similar experiments with the students of four American universities. Altogether 15,863 associations were recorded and classified, and as far as possible the classifications of the American investigators were followed, but it was found that certain rather prominent groups of associations appeared among the Indian students which had not been recorded at all by the American workers, for wherever the total number of any group was less than 100, that group was merged in the general group called "miscellaneous." The total number of associations classified by the American investigators was 14,996, as against 15,863 for Madras, so that we may take the totals as practically equal. Among American students neither political nor religious terms were sufficiently numerous to receive a separate classification; but here there were 1,641 (the largest number) political terms or about 10 per cent of the entire number, and 644 religious terms or about 4 per cent of all. There were 921 associations of the educational type in Madras as against 512 in the American experiments. Vocational terms were much greater here, the number being 875 as against 270. Merchantile terms appeared more frequently among Indian students, the numbers being 424 as compared with 119. Terms expressing kinship occurred 309 times among the Indians; 130 times among the Americans. On the other hand the American results indicate a larger number of associations of the following types: vegetable kingdom (596 as against 410); mineral kingdom (408—166); foods (535—159); interior furnishings (784—290); implements and utensils (758—216); animal kingdom (1202—240); wearing apparel and abrics

(746—147). On the functioning principles of association, particularly contiguity, similarity and contrast, it is normal and natural that the dominant interests come to the fore in processes of association where there is a minimum of control. So it may be taken as giving some indication of the direction of dominant interests, if we can carry on an experiment of this kind sufficiently wide as to give these dominant interests a chance to express themselves normally.

In a similar way one value of the mental test, when it is used to any large degree in this country will be that we shall have thereby an instrument whereby we can tell what are the dominant characteristics of the conscious processes of the people of India. That will furnish us with data with which we can make further comparisons, if we are interested in making comparative studies.

A certain amount of work has been done in comparing different peoples on the basis of mental tests. In the American Army tests a good many negroes were tested, and the conclusion was that the negro was for the most part vastly inferior to the average American, only 16 per cent of them being found to equal or exceed the average for the other citizens. Tests have also shown that the American Indian is not greatly superior to the negro. The children of immigrants from Southern European countries also tested low. But the tests of Chinese and Japanese children showed that their average intelligence was quite as high as the average American child. Mexican children have tested as of an inferior average, but immigrants from Germany, France, England and Scandanavia have all tested to a high average. Terman, in writing in a recent number of *The World's Work*, says that the samplings are not sufficient on which to base too broad generalizations as to racial inferiority or superiority, yet the information received tends in certain definite directions.

We have very little information on which we can base any definite conclusions yet in regard to the mentality, or technically speaking the average intelligence quotient, of Indian subjects as compared with other races. The little that has been done tends to point in the direction of very satisfactory results. But we would like to know not only in a general way, but with mathematical precision what comparisons are legitimate. This is a question that cannot be answered until a sufficient amount of work has been done to furnish us with the necessary data.

There has been a common popular conception that the mentality of women is below that of men, a notion that was abroad in the West, and was even more accentuated in India. But mental tests have completely vindicated the intellectual equality of women with men. The question has been so completely settled for psychology that we do not even see references to it any more. There are other

subsidiary problems, however, that remain to be investigated. One is as to whether there are differences in the special abilities which go to make up the intelligence of the two groups. Is it possible that sex difference may make women stronger in some tendencies, and men stronger in others? Psychology has discovered certain differences that may very well lead in some such direction. We know that the adolescent period dawns earlier in girls than in boys, as also does maturity. Mr. Hoyland's experiments led him to conclude in regard to Indian adolescents that:—

"(i) girls excel boys in altruism and æsthetic interest, but have less regard for truth;

"(ii) girls are more practically-minded than boys, and show more desire to save money; but they are not attracted by a merely domestic career;

"(iii) the girl's maximum period of intellectual development appears to occur at eleven, but the boy's at seventeen."

I have no doubt that experiments with mental tests will compel a revision of Mr. Hoyland's conclusions, particularly the latter. Certainly the ages of eleven for girls and seventeen for boys is too great a disparity for the age of greatest intellectual development. Mention has been made in a previous chapter of the fact that investigators in the West are practically all agreed that the maximum mental development is reached at the age of sixteen. If there is any difference in the ages between India and the West in that regard, it will more than likely be found that the maximum development is reached earlier in India than in the West on account of climatic influences. Such work as has been done in the field is leading workers to think such to be the case, but not enough has been done to reach certain conclusions. The difference in the adolescent period between boys and girls is normally about two years, so that Mr. Hoyland's conclusions of a difference of six years is probably much too great. All that can be said with certainty is that we do not know. But shall we not consider our lack of knowledge an incentive to determine to find out?

Mental tests have come to stay. While appreciating that there are still inaccuracies which we would like to overcome, and that there are traits and abilities which we cannot yet measure satisfactorily, there still remains the fact that there is no device comparable to the mental test for giving us accurate information about mental abilities. If in the West the work is still in its childhood, here in India it has not yet doffed the swaddling clothes of infancy. Wherever educators and psychologists have made use of the method of mental measurement, no matter how much the scepticism with which they began, all have been converted to the reliability of the method for the purpose for which it was devised. It is acknowledged that "individual psychology has achieved its greatest

success in the field of intelligence testing," and indeed that "the developments of the last two decades in this line constitute the most notable event in the history of modern psychology."[1] Certainly we in India do not want to lag behind in making use of the finest technique that scientific psychology has to offer us in measuring the abilities, the achievements and the progress of our future citizens.

---

1 Terman: *Were We Born That Way?* in *The World's Work*, Oct. 1922, p. 655.

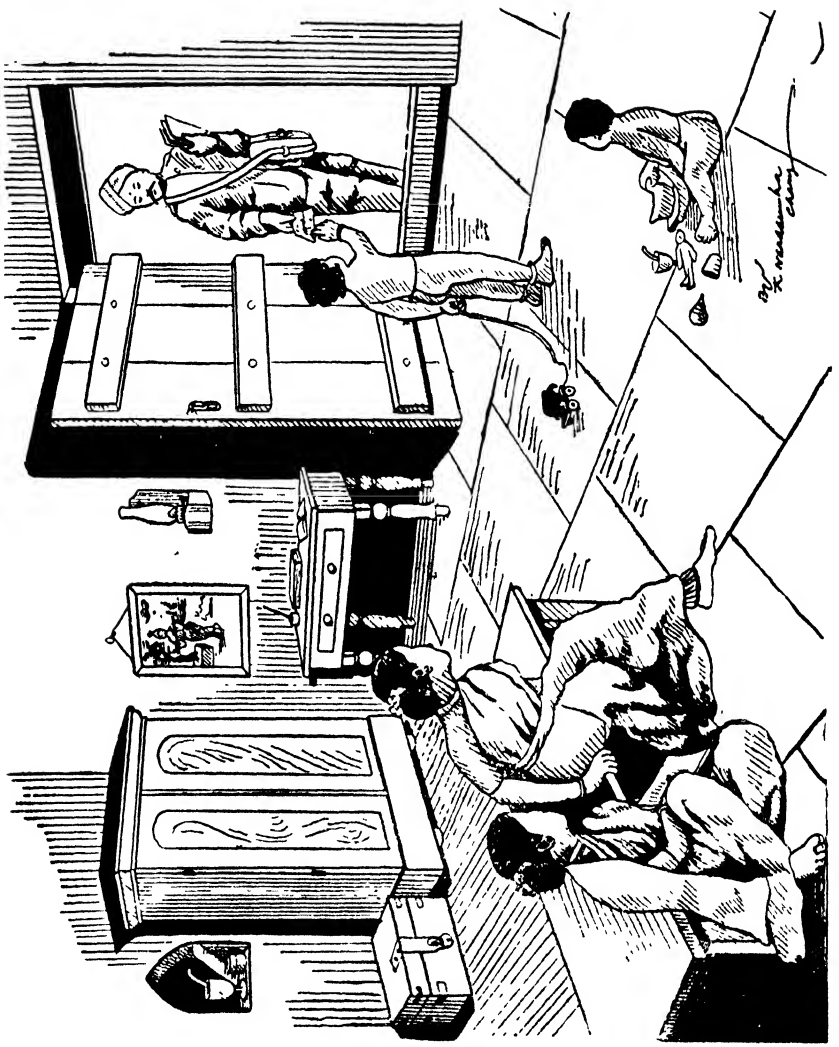


FIG. 1.—AN INDIAN HOME.

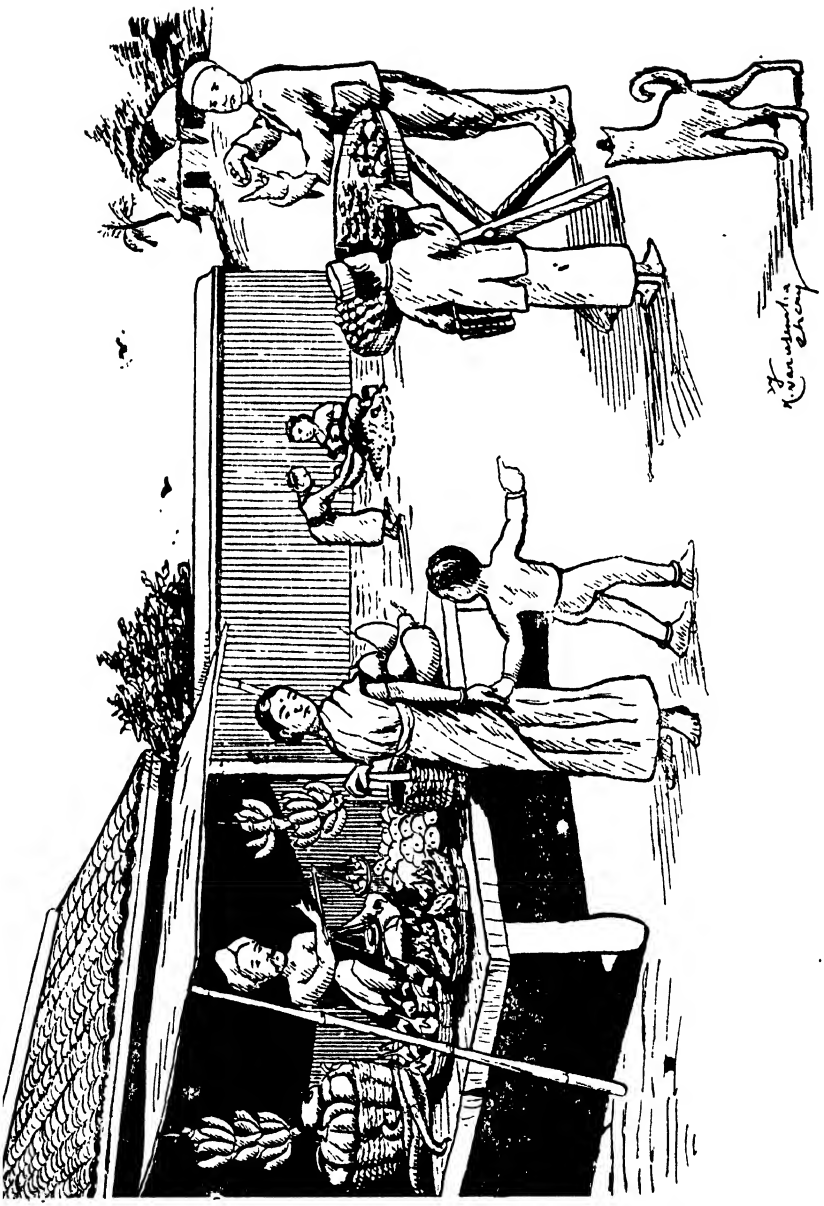


FIG. 2.—THE BAZAAR.

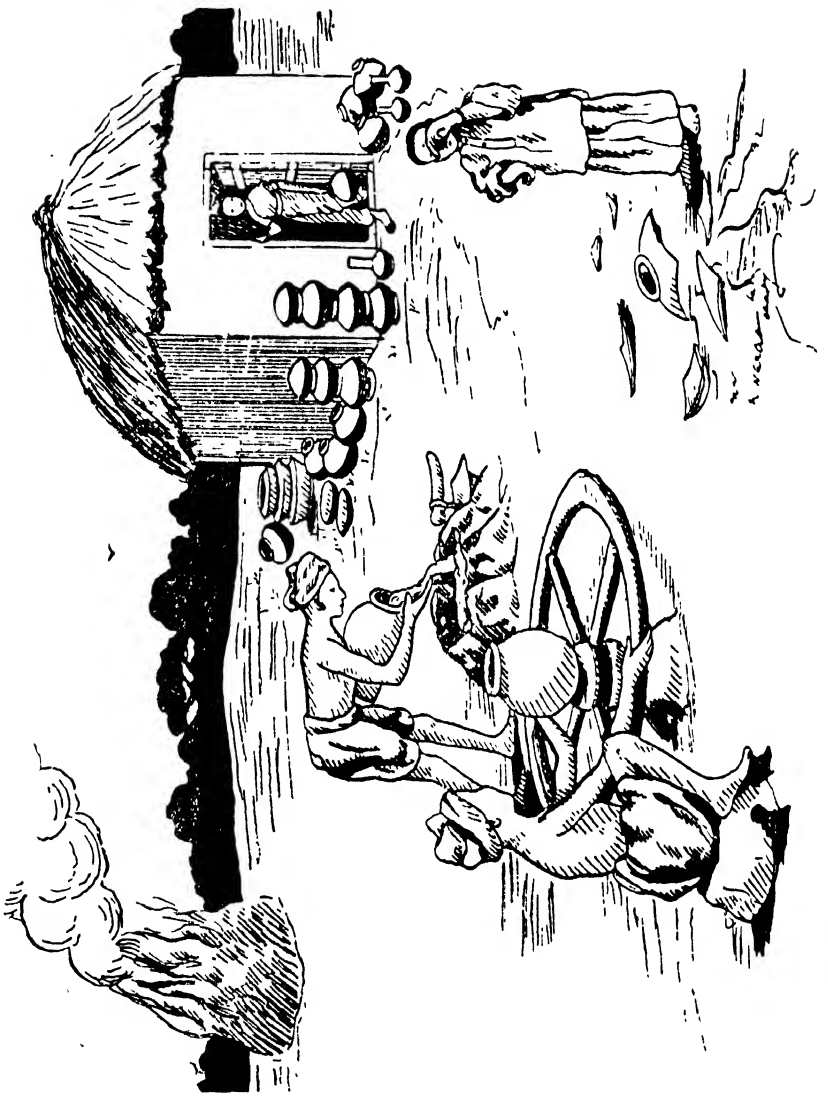


FIG. 3.—THE POTTER.

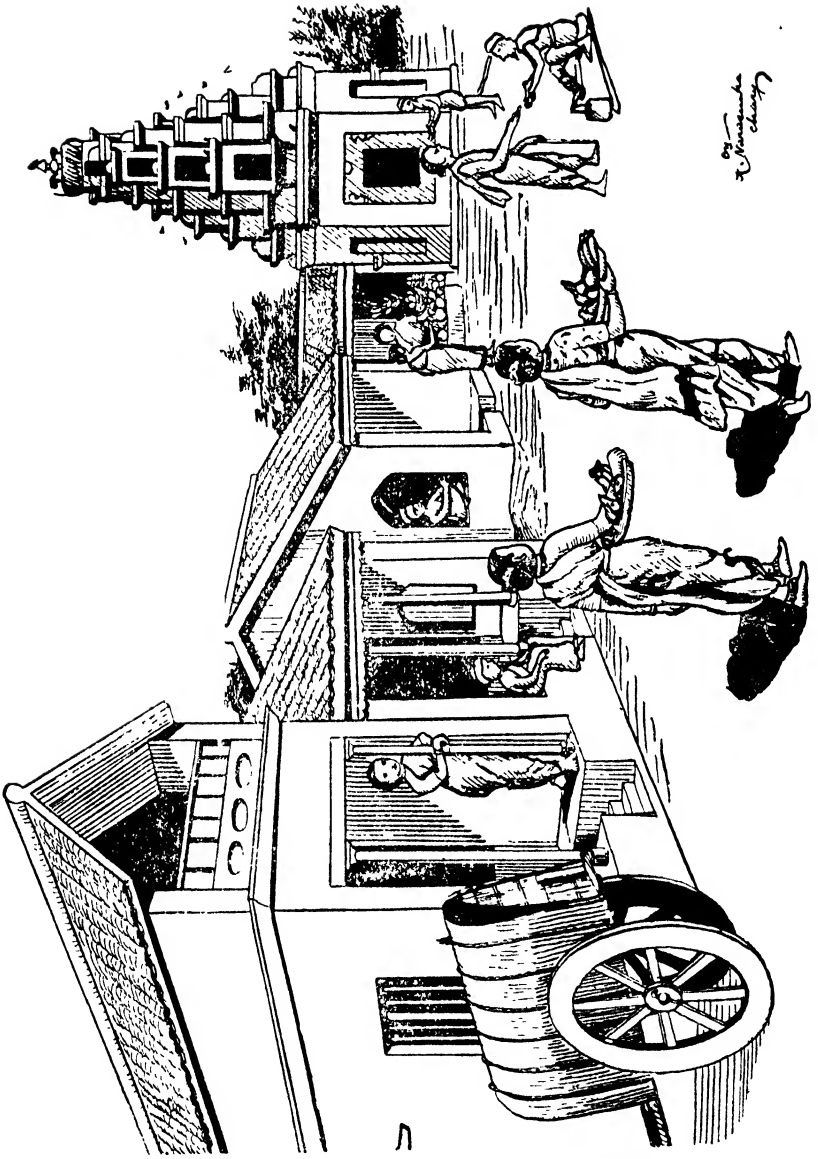


FIG. 4.—A STREET SCENE.

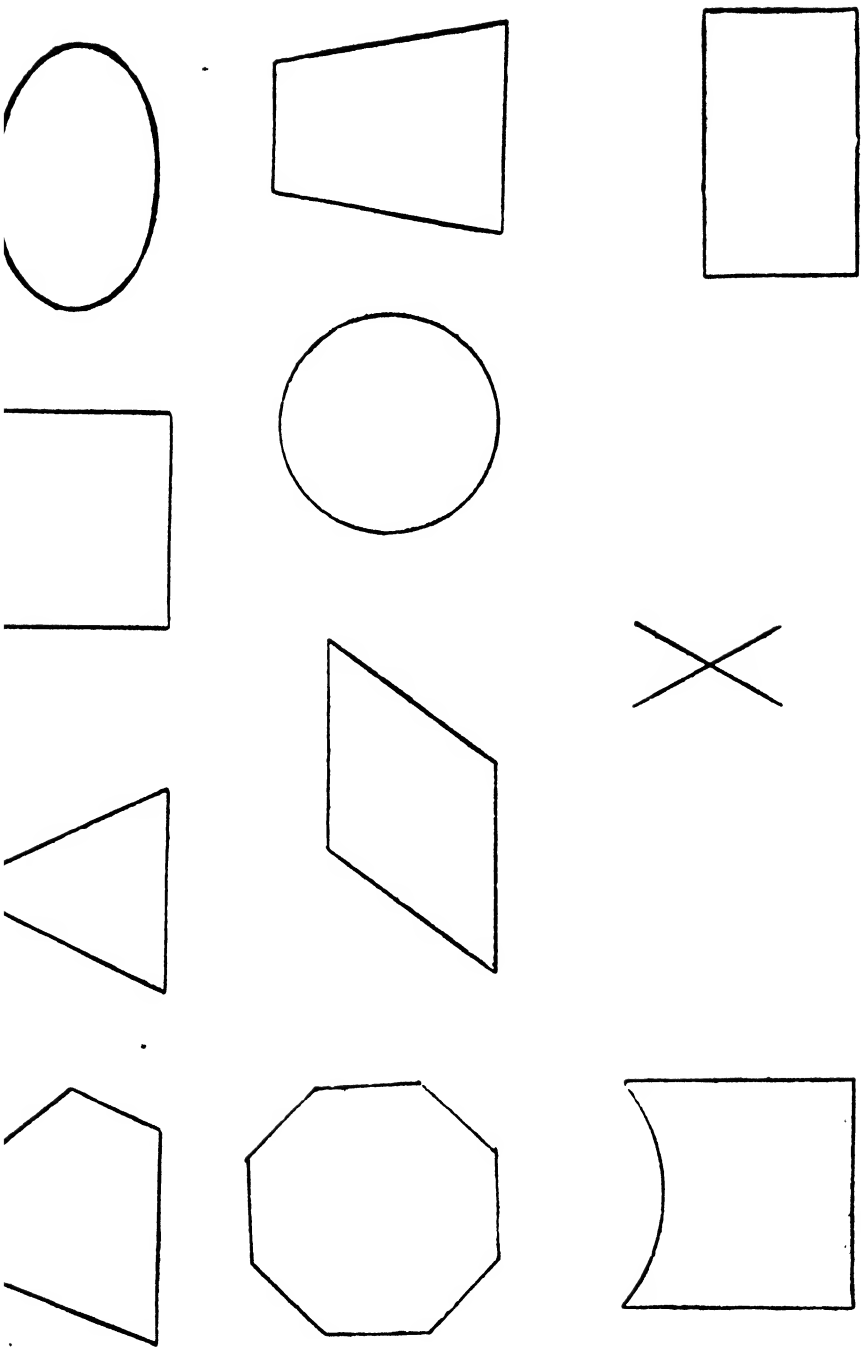


FIG. 5.—DISCRIMINATION OF FORM.

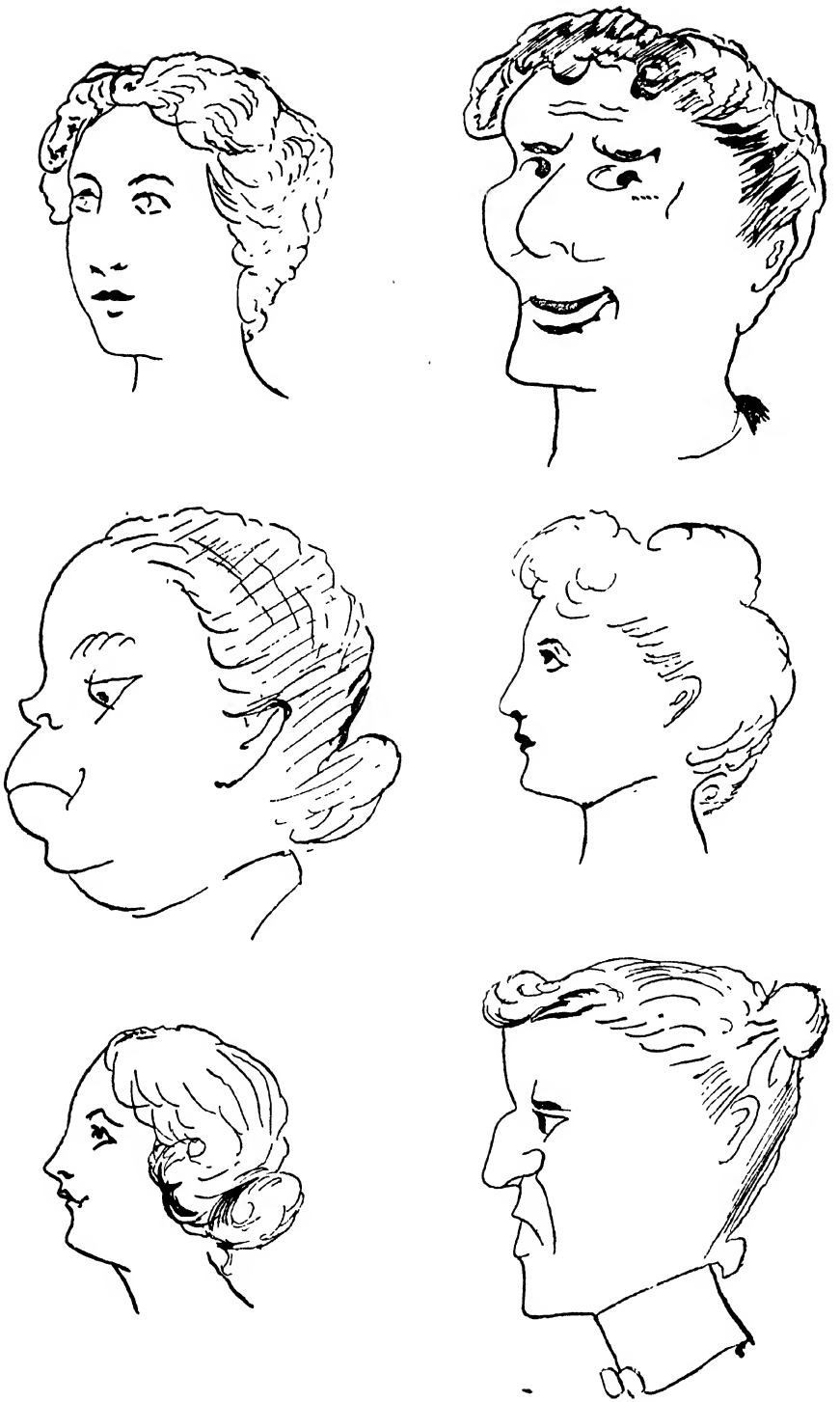


FIG. 5.—COMPARISON OF FACES.

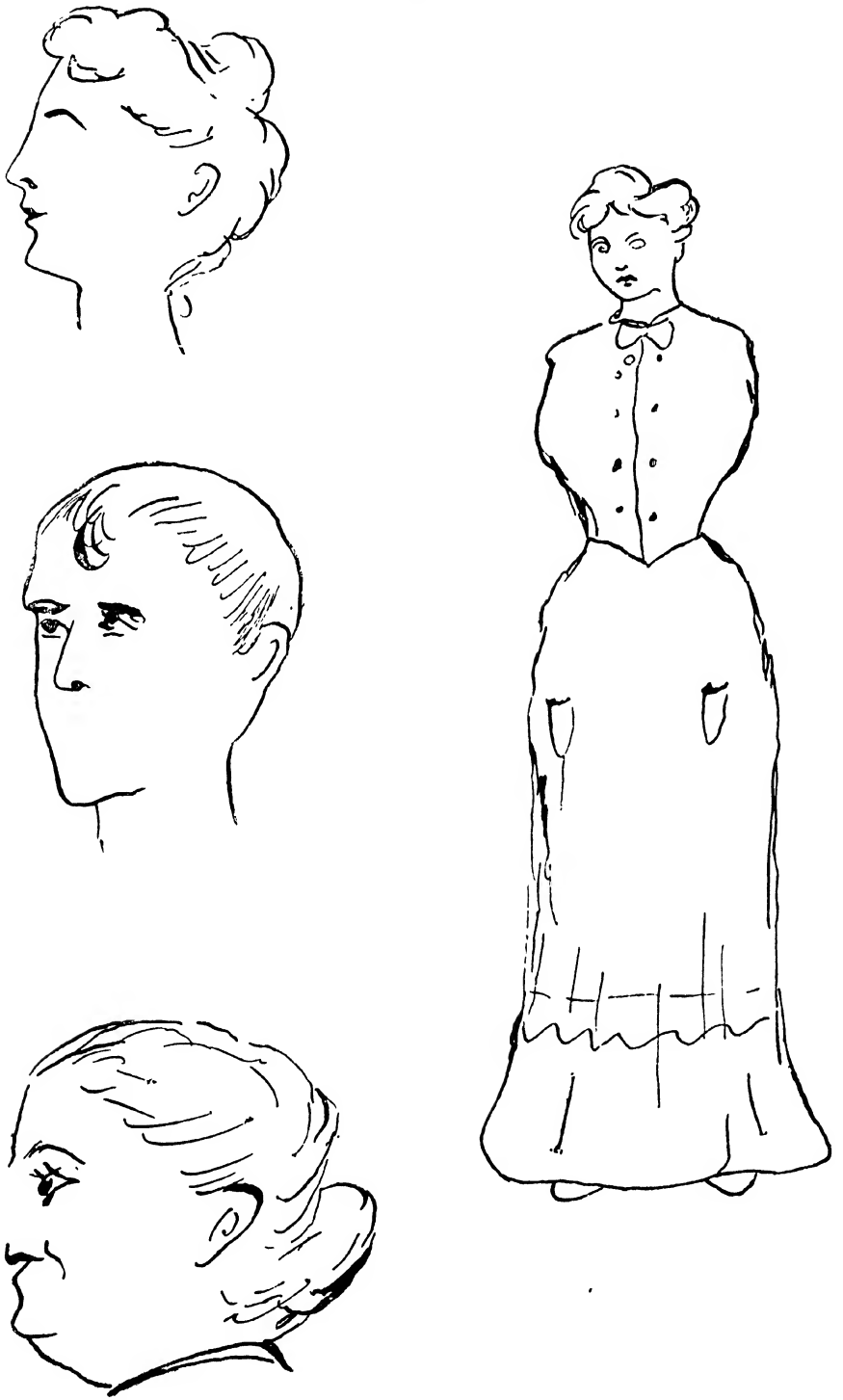


FIG. 7.—FINDING OMISSIONS.

INDEX.

J

K

L

M

N

O

P

V

W

Y